Understanding the Effectiveness of Video Ads: A Measurement Study

S. Shunmuga Krishnan
Akamai Technologies
sarumuga@akamai.com

Ramesh K. Sitaraman
University of Massachusetts, Amherst
& Akamai Technologies
ramesh@cs.umass.edu

ABSTRACT

Online video is the killer application of the Internet. Videos are expected to constitute more than 85% of the traffic on the consumer Internet within the next few years. However, a vexing problem for video providers is how to monetize their online videos. A popular monetization model pursued by many major video providers is inserting ads that play in-stream with the video that is being watched. Our work represents the first rigorous scientific study of the key factors that determine the effectiveness of video ads as measured by their completion and abandonment rates. We collect and analyze a large set of anonymized traces from Akamai's video delivery network consisting of about 65 million unique viewers watching 362 million videos and 257 million ads from 33 video providers around the world. Using novel quasi-experimental techniques, we show that an ad is 18.1% more likely to complete when placed as a mid-roll than as a pre-roll, and 14.3% more likely to complete when placed as pre-roll than as a post-roll. Next, we show that completion rate of an ad decreases with increasing ad length. Further, we show that the ad completion rate is influenced by the video in which the ad is placed. An ad placed in long-form videos such as movies and TV episodes is more likely to complete than the same ad placed in short-form video such as news clips. Our analysis also shows that repeat visitors to a video provider's site complete watching ads at a higher rate than one-time visitors to the site. And, viewers exhibit more patience when watching ads than when faced with performance problems such as a slow-loading video. The abandonment rate for viewers waiting for a slow-loading video to start was more than three times higher than that for viewers watching an ad before their video starts. Finally, we show that about one-third of the viewers who abandon leave in the first quarter of the ad, while about two-thirds leave at the half-way point in the ad.Our work represents a first step towards scientifically understanding video ads and viewer behavior. Such understanding is crucial for the long-term viability of online videos and the future evolution of the Internet.

This is the full version of a paper that appeared in the ACM Internet Measurement Conference (IMC) held from October 23–25, 2013 in Barcelona, Spain.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Performance attributes, Measurement techniques; C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*Client/server*

Keywords

Online videos; Advertisements; Monetization; User behavior; Internet content delivery; Multimedia.

1. INTRODUCTION

Online video is the killer application of the Internet. According to a recent Cisco study more than half of the consumer traffic on the Internet today is related to videos and that fraction is expected to exceed 85% in 2016 [9]. As all forms of traditional media such as news, entertainment and sports migrate to the Internet, video on-demand traffic is expected to triple by 2016 from the levels seen in 2011. Video providers who offer online videos include news channels (such as CNN and Fox News), sports channels (such as ESPN and MLB), movie outlets (such as Hulu and NetFlix), and entertainment providers (such as NBC, ABC, and CBS). Video providers bear the costs of acquiring and delivering the videos to their audience of viewers. Acquisition costs may include production costs for original content or licensing costs and/or revenue sharing for third-party content. The delivery costs often involve contracting with a content delivery service (such as Akamai [18]), who in turn incur the costs for the servers, software, bandwidth, colocation, and power. The runaway success of online videos leaves video providers and the media industry with perhaps their single most vexing problem. How can online videos be monetized? How can they be made viable and profitable?

While successful models for video monetization are still evolving, there are broadly three monetization models that are gaining popularity in the industry. The *subscription-based* model requires users to pay a fee on a periodic basis (usually monthly) to watch videos. The *pay-per-view* model requires users to pay a fee usually on a per-event basis. Finally, a popular model more relevant to our work is the *ad-based* model where viewers do not pay a fee but are shown ads that are placed *in-stream* in the video content.[1]

Driven by both the rapid increase in online video consumption and the intense need to monetize that consump-

[1]We use the term "video" to describe the video being watched, such as a news clip, sports event, or movie. We use the term "ad" to indicate the ad that is played in-stream with the video that is being watched.

tion, it is perhaps not surprising that video ads were the fastest growing category of online ads with spending increasing by about 50% in 2012 [8]. But, how effective are video ads? Are there general causal rules of viewer behavior that govern their effectiveness? What key factors of the ad, of the video, and of the viewer influence an ad's effectiveness? These questions are of great importance to the long-term viability of online videos that are a key part of the Internet ecosystem. However, to our knowledge, they have not been studied in a rigorous scientific fashion, and hence our focus.

1.1 Understanding Ad Effectiveness

The question of how to measure the effectiveness of a video ad is complex. Ads convey a message to the viewer and the key metric for ad effectiveness that is widely used in the media industry is *ad completion rates*. Ad completion rate is the percentage of ads that the viewer watched *completely* without abandoning in the middle. Completion rates are perhaps the most tracked metric in an ad campaign since a viewer watching an ad to completion is more likely to be influenced by it. A related metric is *ad abandonment rate* that measures what fraction of viewers watched what fraction of the ad. The goal in any advertising campaign is to maximize completion rates and minimize abandonment.

In addition to ad completion, there are a few other metrics that are tracked that attempt to measure the response of the viewer after watching the ad. Primary among those is the click-through rate (CTR) that measures the percentage of users who click on a link associated with the ad during or after watching the ad. CTR has the advantage over ad completion rates of capturing an active user response. Though many have argued that CTRs capture only a immediate response but not the long-term impact of the ad that advertiser is hoping to achieve [12].

Another class of metrics for ad effectiveness take the more direct approach of surveying a sample of users who have viewed the ad to determine how much the ad may have increased brand awareness, brand loyalty, and the viewer's intent to buy. While the directness of the approach is an advantage, such surveys are difficult to do at scale and suffer from biases that relate to how the questions are framed and who opts to participate in the survey.

While the video ad industry is yet to evolve a consensus on how to integrate the different ways of measuring ad effectiveness, there is consensus that a basic and important measure is ad completion rate. Thus, we focus on ad completion rate and the associated metric of ad abandonment rate as indicators of ad effectiveness in our study. Our current data set does not currently allow us to measure CTRs or survey responses. But, comparing the different metrics of ad effectiveness is an interesting avenue for future work.

1.2 Our contributions

To our knowledge, our work is the first in-depth scientific study of video ads and their effectiveness. We explore how ad effectiveness as measured by completion rate is impacted by key properties of the ad, of the video, and of the viewer. A key contribution of our work is that we go beyond simple characterization to derive causal rules of viewer behavior that are predictive and more generally applicable. To derive such rules we develop and use a novel technique based on quasi-experimental designs (QEDs).

Our data set is one of the most extensive cross-sections of enterprise videos used in a scientific study of this kind. The data used in our analysis was collected from 33 video providers over a period of 15 days consisting of 362 million videos and 257 million ad impressions that were watched by 65 million unique viewers located across the world.

The metrics that we study such as completion and abandonment rates are critical in the media industry and are widely tracked and reported by ad networks and analytics providers. We expect that the deeper scientific understanding that our work provides for these metrics will have a significant impact on the evolution of monetization models for video. We now list our specific key contributions below.

(1) "Mid-roll" ads placed in the middle of a video had the highest completion rate of 97% while "pre-roll" ads placed in the beginning and "post-roll" ads placed in end yielded drastically smaller completion rates of 74% and 45% respectively. The intuitive reason is that viewers are more engaged with the video during a mid-roll ad causing them to be more patient, while viewers are less engaged in the beginning and at the end of the video. By designing a quasi-experiment, we verify the above intuition by showing that the position of an ad can causally impact its completion rate. We show that an ad is 18.1% more likely to complete when placed as mid-roll than as a pre-roll, and 14.3% more likely to complete when placed as pre-roll than as a post-roll.

(2) 20-second ads had the least completion rate of 60% in our data set, with 15-second and 30-second ads completing at higher rates of 84% and 90% respectively. However, using a quasi-experiment, we show that longer ads are in fact *less* likely to complete. Our causal analysis bolsters our intuition that viewers have less patience to wait for longer ads and would complete fewer of them, provided the other confounding factors are kept similar.

(3) Ads played within long-form video such as TV episodes and movies completed at a higher rate of 87%. While ads played within short-form video such as news clips completed at a lower rate of 67%. A plausible reason is that viewers are more willing to complete an ad that they view as a "cost" if they perceive a greater "benefit" in return from watching the associated video. And, on average, viewers tend to perceive greater benefit from a long-form video than a short-form one. Using a quasi-experiment, we confirm this intuition by showing that an ad is 4.2% more likely to complete if placed in a long-form video than if it is placed in a short-form video, provided all other factors are similar. Note that the magnitude of impact of video length on ad completion rates is smaller when confounding factors are accounted for than what a simple correlation suggests (4.2% versus 20%).

(4) Using information gain ratios, we show that the contents of the video and the ad have high relevance for completion rates, while the connectivity of the viewer had the lowest relevance.

(5) Industry folklore suggests that viewers are less likely to abandon ads when watching them in the evenings or weekends when they tend to be more relaxed and have more spare time. However, we did not find any supporting evidence in our data as we did not observe a significant influence of either time-of-day or day-of-week on ad completion rates.

(6) Repeat visitors to a video provider's site had noticeably higher ad completion rates than one-time visitors to that site. In particular, repeat visitors had a completion rate of 84.9% and one-time visitors only 78%. Further, repeat visitors had higher ad completion rates for all major subcategories of ads, such as pre-rolls, mid-rolls, and post-rolls. A plausible reason for this phenomena is that repeat visitors to a site are more interested in the video content offered at the site and are hence more likely to complete ads on that site.

(7) Viewers are more impatient in waiting for a slow-loading video to start than for a pre-roll ad to finish. For instance, at the 10-second mark, 45.8% of the viewers waiting for a slow-loading video have abandoned, compared to only 13.4% of viewers watching a pre-roll ad. Waiting for an ad to complete is less frustrating since it is an expected wait of a known duration. While slow-loading videos and other performance issues are usually unexpected and are of unknown duration, and hence they more frustrating to the viewer leading to greater abandonment. This result quantitatively suggests that if the media player can predict that a video will start up late due to an underlying network problem, one can likely reduce viewer abandonment by inserting an ad of the appropriate length instead of showing a blank screen or a "spinning wheel".

(8) In our study of ad abandonment rates, we show that a significant set of viewers abandon soon after the ad starts. The abandonment rate is initially higher but tapers off over time as the ad plays. About one-third of the viewers who eventually abandon leave in the first quarter of the ad, while about two-thirds leave at the half-way point in the ad.

2. BACKGROUND

2.1 The Video Ads Ecosystem

The video ad ecosystem consists of four types of entities. *Video providers* own and manage video content, e.g., NBC, CBS, CNN, Hulu, Fox News, etc. Advertisers offer ads that can be played in-stream with the video. Ad insertion is managed by an ad delivery network such as Freewheel [5], Adobe Auditude [1], or Video Plaza [7]. The ad network brings together the video providers (i.e., publishers) who offer videos and the advertisers who offer ads. An ad network has an ad decision component that decides what ads to play with which videos and where to position those ads. Both the ads from the advertisers and the videos from the video providers need to be streamed to the viewer with high performance. For that reason, both the ads and videos are typically delivered to the viewers using content delivery networks (CDNs) such as Akamai [10, 18]. Thus, CDNs are cognizant of both the video content and ads embedded within them. The mechanism for inserting the ad is commonly performed by the user's media player when it is playing the video. When it is time to play an ad, the media player redirects to the ad network that chooses the ad and plays it within an ad player.When the ad completes the control returns to the user's media player that continues to play the video content.

2.2 Views, viewers, and visits

We describe a user watching a video with ads, defining terms along the way that we will use in this paper.

Viewer. A viewer is a user who watches one or more videos using a specific media player installed on the user's device. A viewer is uniquely identified and distinguished from other viewers by using a GUID (Globally Unique Identifier) value that is set as a cookie when the media player is accessed. To identify the viewer uniquely, the GUID[2] value is generated to be distinct from other prior values in use.

Views. A view represents an attempt by a viewer to watch a specific video. A typical view would start with the viewer initiating the video playback, for instance, by clicking the play button of the media player[3] (see Figure 1). During a view, the media player might first play an ad called a *pre-roll* before the actual video content begins. (We use the term *video* to denote the actual video that the viewer wants to watch to distinguish it from the *ad* that the viewer is also shown.) Further, the video may be interrupted in the middle with one or more ads called *mid-rolls.* Finally, an ad might be shown when the video ends called a *post-roll.* Each showing of an ad, whether or not it is watched completely, is called an *ad impression. Ad completion rate* is the percent of ad impressions that were played to completion[4].

Typically, viewers do not have the ability to "skip" ads and must either watch the ad in order to watch the video content that follows, or just abandon the ad and the view all together. Our data sets have such non-skippable ads that is the standard for enterprise video. Recently YouTube that has a large fraction of user-generated videos has been experimenting with pre-roll ads that have a mandatory non-skippable part that must be viewed but can be skipped beyond that point. But, it is not yet common within enterprise videos and is not represented in our data set.

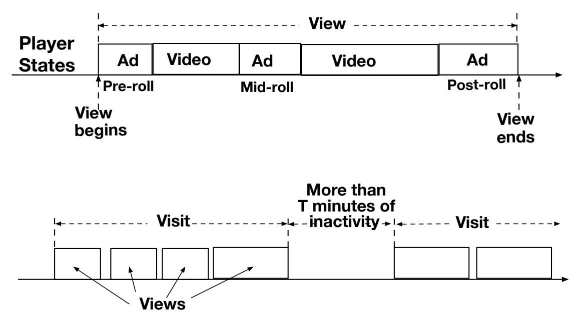

Figure 1: Views and Visits

Visits. A visit is intended to capture a single session of a viewer visiting a content provider's site to view videos. A visit is a maximal set of contiguous views from a viewer at a specific video provider site such that each visit is separated from the next visit by at least T minutes of inactivity, where we choose $T = 30$ minutes[5](see Figure 1).

[2]In most implementations, the GUID is tied to the device or the desktop of the viewer. Thus, we cannot always detect cases where one user watches video on another user's device.
[3]In other cases, a view may be initiated automatically using a play list or by other means.
[4]If an ad is played completely it is *likely* that it was watched completely by the viewer. However, we are not able to measure whether or not a viewer is actually watching the ad or if he/she has shifted focus elsewhere when the ad is playing.
[5]Our definition is similar to the standard notion of a visit (also called a session) in web analytics where each visit is a set of page views separated by a period of idleness of at least 30 minutes (say) from the next visit.

Type	Factor	Description
Ad	Content	defined by unique name
	Position	Pre-, mid-, post-roll
	Length	15-, 20-, and 30-second
Video	Content	defined by unique url
	Length	Short-form, Long-form
	Provider	News, Movie, Sports, Entertainment
Viewer	Identity	defined by unique GUID
	Geography	Country and Continent
	Connection Type	Mobile, DSL, Cable, Fiber
	Temporal	Time of day, Day of week
	Frequency	Frequency of visits to a site

Table 1: Potential factors that relate to or impact viewer behavior and ad completion.

2.3 Potential factors that impact ad completion

We study three sets of key factors that potentially influence ad completion that we list in Table 1 and discuss below.

Ad-related factors. The actual ad and its contents as identified by its unique name is a first factor. Ad position relates where it was inserted in a video view and can either be pre-roll, mid-roll, or post-roll. The most common ad lengths in our study are 15-second, 20-second, and 30-second ads.

Video-related factors. The first factor is the video content itself as identified uniquely by its url[6]. Besides the actual content of the video itself, we isolate two important factors. The *video length* can be used to differentiate short-form from long-form videos. The IAB (Interactive Advertising Bureau) which is a major industry body for online video advertising defines long-form video as videos lasting over 10 minutes and short-form as those under 10 minutes [6]. We adopt this standard definition in our work. Typically, short-form and long-form videos are qualitatively different. Short-form video are usually smaller clips for news, weather, etc. Long-form videos are typically TV episodes, movies, sports events, etc. Most long-form videos possess a "content arc" with a beginning, middle and end.

Viewer-related factors. A viewer can be uniquely and anonymously identified by their GUID. Besides their identity, we consider three important attributes of the viewer. The geographical location of the viewer at the country level encapsulates several social, economic, religious, and cultural aspects that could influence his/her viewing behavior. In addition, the manner in which a viewer connects to the Internet, both the device used and typical connectivity characteristics can influence viewer behavior. The primary connection types are mobile, DSL, cable, and fiber (such as AT&T's Uverse and Verizon's FiOS). Further, it is plausible that the time-of-day and the day-of-week[7] when the ad was watched could potentially influence its completion rate. For instance, folklore holds that people have more time in the weekend and evenings, leading them to be more relaxed and more patient with video ads. However, as we show in Section 5.3.3, we did not observe a significant influence of either time-of-day or the day-of-week on ad completion rates. Finally, it is possible that the frequency with which a viewer visits a site could relate to ad completion rates, as viewers who visit more frequently may likely have more interest in the videos offered at that site. For instance, a football fan visiting a sports site frequently to watch highlights might be more eager to complete watching the ads on that site.

It is worth noting that many of the video and viewer related factors considered here are known to significantly impact viewer behavior in the context of viewer tolerance to performance degradation from our prior work [14]. So, it is natural to consider these in the different behavioral context of viewer tolerance to ads. The ad-related factors considered here, besides being natural to consider, are widely tracked in industry benchmarks.

3. DATA SETS

The data sets that we use for our analysis are collected from a large cross section of actual users around the world who play videos using media players that incorporate the widely-deployed Akamai's client-side media analytics plug-in [2]. When video providers build their media player, they can choose to incorporate the plugin that provides an accurate means for measuring a variety of video and ad metrics. When the viewer uses the media player to play a video, the plugin is loaded at the client-side and it "listens" and records a variety of events that can then be used to stitch together an accurate picture of exactly what the viewer experienced and what the viewer did.

When a view is initiated, say with the user clicking the play button, metrics such as the time when the view was initiated, the video url that uniquely identifies the content, video length, whether it is live or on-demand, the video provider, the amount of video watched, the bitrate(s) at which it was streamed, and several other detailed characteristics pertaining to the video are recorded. Likewise, when an ad is inserted, ad-related metrics such at what point in the video the ad was inserted, the ad name that uniquely identifies the content of the ad, the ad length, the amount of the ad that was actually played, and whether the ad completed or not are recorded. Further, detailed information about the viewer is recorded including the GUID that uniquely identifies the viewer, current ip address, network, geography, and connection type. Once the metrics are captured by the plugin, the information is "beaconed" to an analytics backend that we use to process the huge volumes of data. From every media player at the beginning and end of every view, the relevant measurements are sent to the analytics backend. Further, incremental updates are sent at a configurable periodicity as the video is playing, typically once every 300 seconds. All relevant fields in the data set used in our study are measured and stored in an *anonymized* fashion so as to not include any PII or sensitive information.

3.1 Data Characteristics

The Akamai CDN serves a significant fraction of world's online videos and ads. We selected a large, characteristic cross-section of 33 video providers including news sites, sports sites, movie providers, and entertainment channels who have an ad-based monetization model. We tracked all videos and ads for these providers over a period of 15 days

[6]Note that if two different providers are showing the same movie with different urls, we consider them different videos. Detecting them to be the same content is intrinsically very difficult as there is no universally accepted naming system across video providers.

[7]Time-of-day and day-of-week is computed using the *local* time for the viewer based on his/her geographical location.

in April 2013. About 94% of the video views were for on-demand content and the rest were live events. We only consider on-demand videos that currently form the bulk of the videos for our study.

Our data is among the most extensive ever studied for video ads and consisted of 257 million ad impressions that were watched by over 65 million unique viewers located in all major continents of the world. On average, viewers spent 8.8% of their time watching ads as opposed videos. Table 2 summarizes some basic statistics of our data. The geography

	Total	Per View	Per Visit	Per Viewer
Views	362 mil	N/A	1.3	5.6
Ad Impressions	257 mill	0.71	0.92	3.95
Video Play (in minutes)	777 mil	2.15	2.79	11.96
Ad Play (in minutes)	75 mil	0.21	0.27	1.15

Table 2: Key statistics of our data set.

of the viewer was mostly concentrated in North America, and Europe that together originate the bulk of video traffic. One continent where we could not obtain proportional representation is Asia that accounts for significant video traffic but where many video providers do not yet have the software changes required in their media player to track ads. The connection types were dominated by cable, though the other categories have a solid representation as well (cf. Table 3).

Viewer Geography	Percent Views	Connection Type	Percent Views
North America	65.56%	Fiber	17.14%
Europe	29.72%	Cable	56.95%
Asia	1.95%	DSL	19.78%
Other	2.77%	Mobile	6.05%

Table 3: Geography and connection type.

The ad length distribution is shown in Figure 2. The ad lengths clustered around 15-, 20- and 30-second marks and were clustered into those categories respectively. The distribution of the video lengths for short-form and long-form videos are shown in Figure 3. The mean length of a short-form video is 2.9 minutes and that of long-form video is 30.7 minutes. The most popular duration for long-form video was 30 minutes that is typical for a TV episode.

4. ANALYSIS TECHNIQUES

We seek to understand how a set of key factors such as those in Table 1 impact viewer behavior metrics such as ad completion and abandonment rates. We use *correlational tools* such as Kendall correlation and information gain described in Section 4.1 to characterize the observed data. While correlational analysis is important as a description of *what is*, they don't necessarily have the ability to predict *what will be*. The ability to predict often requires a deeper inference of a *causal* rule between the key factor and the

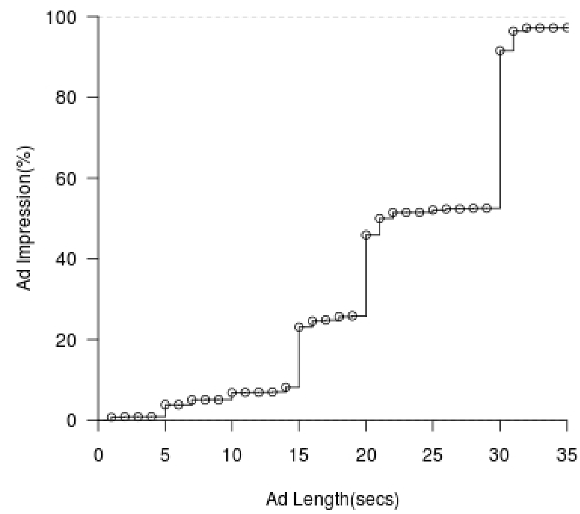

Figure 2: CDF of ad length showing the three major clusters at 15-, 20- and 30-seconds.

viewer behavior metric. To provide an example, a correlational analysis of the observed data will be able to say that mid-roll ads have a higher completion rate than pre-roll ads in the observed data. However, this fact does necessarily imply a causal rule that states that placing an ad as mid-roll will likely cause higher completions than placing the same ad as a pre-roll. The value of causal inference over a purely correlational one is that it extracts general rules of viewer behavior from the data that can be applied to a more general or even different context. In our prior work [14], we introduced an innovative tool called Quasi Experimental Design (QED) that we adapted from the social and medical sciences for use in network measurement research. In this work, we take the next step and further evolve this technique to extract causal rules pertaining to video ads. We describe our technique in Section 4.2.

4.1 Correlational Analysis

To study the impact of a key factor X (say, ad length) with a viewer behavioral metric Y (say, completion rate), we start by visually plotting factor X versus metric Y in the observed data. Then, when relevant, we compute Kendall's correlation coefficient τ that takes values in the interval $[-1, 1]$ where τ near 1 means that larger values of X are associated with larger values for Y, τ near -1 means that larger values of X are associated with smaller values of Y, and τ near 0 means that X and Y are independent.

A key technique that we use to quantify the influence of factor X on metric Y is the *information gain ratio* [13]. Information gain ratio measures the extent to which the variability of Y is reduced by knowing X. That is, information gain is the entropy of Y (denoted by $H(Y)$) minus the entropy of Y given X (denoted by $H(Y|X)$). Normalizing the information gain, we obtain the information gain ratio de-

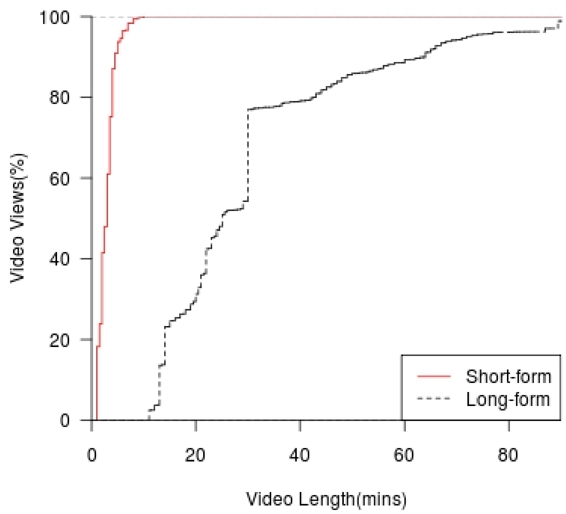

Figure 3: CDF of video length for long-form and short-form videos.

noted by

$$IGR(Y,X) = \frac{H(Y) - H(Y|X)}{H(Y)} \times 100.$$

It is instructive to view the two extreme cases. Suppose knowing X perfectly predicts Y, then $H(Y|X)$ is zero since there is no variability left in Y and $IGR(X,Y)$ is 100%. In the other extreme, suppose that X and Y are statistically independent. In that case, $H(Y|X)$ simply equals $H(Y)$ since knowing X says nothing about Y and $IGR(Y,X)$ is 0%. In all our results, IGR is somewhere in between and is a quantitative indicator of the extent of a factor's influence on a viewer behavioral outcome.

4.2 Causal Analysis using QEDs

A correlational analysis of factor X (say, ad length) and a viewer behavior metric Y (say, completion rate) could show that X and Y are associated with each other. A primary threat to a causal conclusion that an *independent variable* X causes the *dependent variable* Y is the existence of *confounding variables* that can impact both X and Y. To take an example that we describe in greater detail in Section 5.1.3, suppose we want to infer a causal rule that a longer ad causes viewers to complete the ad less often. Simply correlating completion rate and ad length is not sufficient. In fact, 20-second ads complete less often than 30-second ads in the observed data, apparently violating the rule. To derive a causal conclusion, one would need to account for the confounding factor of ad position, since 30-second ads are often placed as mid-rolls and as we show mid-rolls have a higher completion rate independent of length.

A primary technique for showing that an independent variable X (also called the treatment variable) has a causal impact on a dependent variable Y (called the outcome variable) is to design a *quasi-experiment*. Quasi-experiments were developed by social and medical scientists and has more than 150 years of history in those domains [20]. In particular, we use a specific type of QED called the matched design [19] where a treated individual (in our case, a view or viewer) is randomly matched with an untreated individual, where both individuals have similar values for the confounding variables. Consequently, any difference in the outcome for this pair can be attributed to the treatment. By creating a large collection of matched pairs and assessing the differential outcome of the paired individuals, one can isolate the causal impact of X on Y.

Adapting QEDs to our situation, our population typically consists of views. The independent variable is one of the factors in Table 1 (say, ad position). The treated and untreated sets have two different values of the independent variable that we want to compare (say, mid-roll versus pre-roll). Our outcome is a function of the behavioral metric under study, such as ad completion. The confounding factors that need to be matched so that they have similar values are typically other key factors in Table 1 except the independent variable that is varied, since the other factors could confound the outcome. We form comparison sets by randomly matching each treated view with an untreated view such that they have similar values for the confounding variables and differ only in the independent variable. For instance, to study the impact of ad position, we match views that belong to similar viewers watching the same ad within the same video, neutralizing the impact of the confounding variables. By forming a large number of such pairs and by studying the behavioral outcomes of matched pairs one can deduce whether or not the treatment variable X has a causal effect on variable Y, with the influence of the confounding variables neutralized.

Statistical Significance of the QED Analysis.

As with any statistical analysis, it is important to evaluate whether the results are *statistically significant* or if they could have occurred by random chance. As is customary in hypothesis testing [16], we compute the p-value which evaluates the probability that the observed outcome from a QED happened by chance, assuming that the null hypothesis holds. The null hypothesis states that there is no impact of the treatment on the outcome. To evaluate the p-value we use the sign test that is a non-parametric test that makes no distributional assumptions and is particularly well-suited for evaluating matched pairs in a QED setting [21]. A low p-value means that our results are statistically significant. The choice of the threshold is somewhat arbitrary and in medical sciences a treatment is considered effective if the p-value is at most 0.05. We can achieve much higher levels of significance owing to the large numbers of treated-untreated pairs in our QEDs (order of 100,000) in relation to what is typical in the medical sciences (in the 100's). However, our results are unambiguously significant and not very sensitive to the choice of significance level. We refer to our prior work [14] for a more detailed treatment of QEDs.

Some Caveats.

It is important to understand the limitations of our QED tools, or for that matter *any* experimental technique of inference. Care should be taken in designing the quasi-experiment to ensure that the major confounding variables are explicitly or implicitly captured in the analysis. If there exists confounding variables that are not easily measurable (example, the gender of the viewer) and/or are not identified and con-

Type	Factor	IGR
Ad	Content	32.29%
	Position	15.1%
	Length	12.79%
Video	Content	23.92%
	Length	18.24%
	Provider	15.24%
Viewer	Identity	59.2%
	Geography	9.57%
	Connection Type	1.82%

Table 4: Information gain ratio (IGR) for ad completion rate.

trolled, these unaccounted dimensions could pose a risk to a causal conclusion, but only if they turn out to be significant. *Our work on deriving a causal relationship by systematically accounting for the confounding variables must not be viewed as a definitive proof of causality, as indeed there can be no definitive proof of causality. But, rather, our work increases the confidence in a causal conclusion by accounting for potential major sources of confounding.* This is of course a general caveat that holds for all domains across the sciences that attempt to infer causality from observational data.

5. AD COMPLETION RATE

We study the ad completion rate metric that is a key measure of ad effectiveness. The completion rate can be influenced by the characteristics of three entities that we examine in turn: the ads themselves, the videos that the ads are embedded in, and the viewer who is watching the videos and ads. We evaluate the relevance of each of these factors to the completion rate by computing their information gain ratio shown in Table 4. Enormous effort goes into creating the ad and video content to make it as captivating as possible. It is interesting that both show high information gain, perhaps indicating that content does matter.The information gain ratio of the viewer is very high. This is at least in part due to the fact that 51% of the viewers watched only one ad resulting in either a 0% or 100% completion rate. In those cases, knowing the viewer perfectly predicts the completion rate. Information gain is known to be counter-intuitive for factors like viewer that can take millions of values each with small individual weights. The information gain from connection type was the least, as viewers showed lesser variations in their patience for completing ads across the different connection types. This is in contrast with our earlier work on viewer patience in the context of video performance [14] where viewers with worse connectivity had more patience for a video to start up.

5.1 Impact of Ad-related Factors

We examine three factors that relate to the ad itself: the ad's content as identified by its unique name, the position in which the ad was played, and the length of the ad.

5.1.1 Ad Content

For each unique ad, we can define its completion rate as simply the fraction of ad impressions where the ad was watched to completion by the viewer. We plot the percent of ad impressions (y-axis) attributed to ads with completion rate smaller than some x-value (cf. Figure 4). We can see

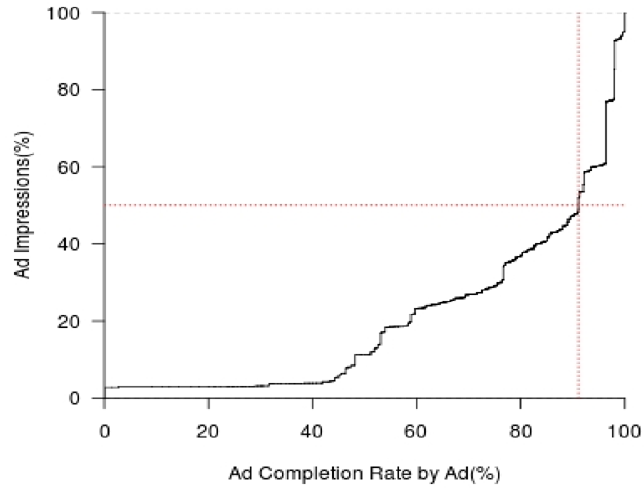

Figure 4: The percent of ad impressions y attributed to ads with ad completion rate smaller than x. 50% of the ad impressions are from ads with completion rate at most 91%.

from the figure that ads complete at varying rates with some having low completion rates with others completing 90+% of the time. Further, 25% the ad impressions come from ads with completion rate under 66%, and 50% come from ads with completion rate under 91%.

5.1.2 Ad Position

We analyze the impact of ad position on the likelihood that a viewer watches the ad to completion. We first take a simple correlational approach of categorizing the position in which the ad was played and computing the completion rates for each category. Our analysis shows that mid-roll ads completed most often, followed by pre-roll and post-roll ads (cf. Figure 5).

Assessing Causal Impact.

Our observational results support the intuition that ads placed in the middle of the content have the most likelihood of being watched, since the viewer is engaged with the content when the ad is shown, wants to watch the rest of the video, and is thus more willing to tolerate the ad. Whereas ads placed as pre-roll run a greater risk of viewers abandoning and going elsewhere, since they have not yet started watching their content and hence are not yet engaged with it. Further, an ad placed at the end of the content as a post-roll runs an even greater risk of viewers leaving since the content that they wanted to watch has completed, and so they are less motivated to sit through an ad. Based on our observational results, we assert the following causal rule.

RULE 5.1. *On average, a viewer is more likely to complete watching an ad that is placed as a mid-roll than when the same ad is placed as a pre-roll. In turn, a viewer is more likely to complete watching an ad that is placed as a pre-roll than when the same ad is placed as a post-roll.*

Note that the correlational analysis in Figure 5 is not sufficient to show that the rule holds, as there are potential con-

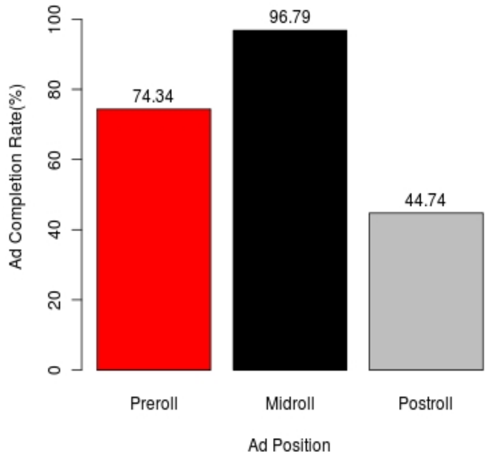

Figure 5: Mid-roll ads complete most often as the viewer is already engaged by the video and wants to watch more.

founding factors such as the ad length, video length, content provider, viewer geography, and viewer connectivity that can negate such an assertion. For instance, the following plausible scenario could be threat to our asserted rule. It *could* be possible that mid-roll ads appear largely in longer content such as TV episodes and movies, and perhaps ads placed in longer content have higher completion rates than ads placed in shorter content *irrespective* of their position. Thus, mid-roll ads could have a higher completion, not by virtue of them being placed in the middle of the content, but simply by being more likely to be placed in longer content.

QED. To carefully assess the impact of ad position in isolation by accounting for other potential confounding factors, we design a quasi-experiment as described by the matching algorithm in Figure 6. To compare the effect of placing an ad as mid-roll versus placing as pre-roll, the algorithm finds matched views (u, v) from two similar viewers who have the same connection type and geography. Further, the two matched views are for exactly the same video and the same ad. The primary difference between the matched views is that the ad was played in different positions, i.e., one view u has the ad as a mid-roll while the other view v has the same ad as a pre-roll. Note that a positive value for net outcome provides positive (supporting) evidence for the rule that an ad in mid-roll completes more often than the same ad as pre-roll, while a negative value provides negative evidence for the asserted rule. The algorithm in Figure 6 can be used to compare any pair of ad positions with minor modifications. For instance, to compare pre-roll with post-roll, we can apply the same algorithm with pre-roll as the treated set T and post-roll as the untreated set C.

QED Results. The results for the quasi-experiment are shown in Table 5. These results show that on average ads run in the mid-roll position are 18.1% percent more likely to complete than the same ad run in the pre-roll position for the same video content for a similar viewer. Further, ads run in the pre-roll position are 14.3% percent more likely to complete than the same ad run in the post-roll position for the same video content for a similar viewer. The results confirm the causal impact of ad position on completion rates and establish Rule 5.1 in a causal and quantitative manner. Further, using the sign test, the p-value for each quasi-experiment was at most 1.98×10^{-323}, confirming the statistical significance of the results.

Matching Algorithm

Matched: similar viewer, same ad, same video.
Independent: ad position.

1. *Match step.* Let the treated set T be the set of all views that had a mid-roll ad and let the untreated set C be the set of all views that had a pre-roll ad. For each $u \in T$ that had some ad α as mid-roll, we pick uniformly and randomly from the set of candidate views $v \in C$ such that u and v belong to similar viewers with the same geography and connection type who are watching the same video and the same ad α, except that the ad α was played as mid-roll in u but played as pre-roll in v[a]. The matched set of pairs $M \subseteq T \times C$ have the same or similar attributes for the confounding variables that are matched and differ only on the independent variable.

2. *Score step.* For each pair $(u, v) \in M$, we compute an $outcome(u, v)$ to be $+1$ if the matched ad was completed in u but not completed in v, - 1 if the matched ad was completed in v but not in u, and 0 otherwise. Now,

$$Net\ Outcome = \frac{\sum_{(u,v)\in M} outcome(u, v)}{|M|} \times 100.$$

[a]If no match v exists for a u, then no pair is formed.

Figure 6: The matching algorithm that compares ads placed as mid-roll (treated) versus pre-roll (untreated) while accounting for the other confounding variables such as the ad, video, and viewer characteristics.

Treated/Untreated	Net Outcome
mid-roll/pre-roll	18.1%
pre-roll/post-roll	14.3%

Table 5: Net QED outcomes support the rule that placing an ad as a mid-roll can cause greater completions than as a pre-roll or as a post-roll.

Discussion.

(1) Note that the impact of ad position on ad completion rates turns out to be smaller (but still significant) when the confounding factors are accounted for than in the simpler correlational analysis of Figure 5.
(2) If mid-rolls are so effective, why not place only mid-roll ads? While our results show that positioning an ad as mid-roll increases its likelihood of completion, it is not a recommendation for advertisers to place only mid-roll ads. If an ad network wants to achieve a certain number of completed ad impressions one needs to worry about both the audience size and the ad completion rate. Audience size for pre-roll

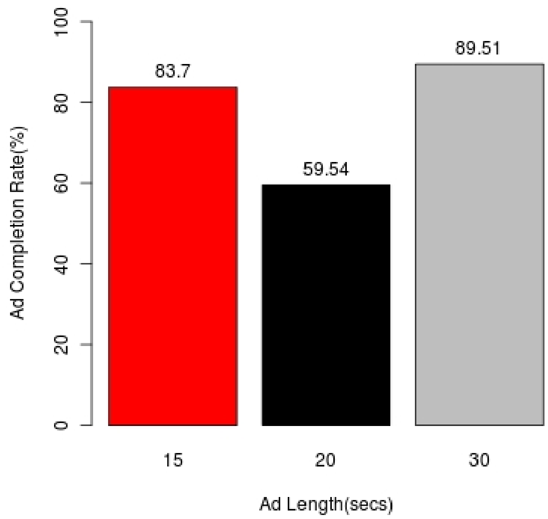

Figure 7: The measured ad completion rates in our data set did not decrease with ad length as expected. The 30-second ad, while longer, had the highest completion rate in part due to being placed more frequently in the mid-roll position.

ads are larger than mid-roll ads simply because viewers drop off before the video progresses to a point where a mid-roll ad can be played. Likewise, the audience size of a mid-roll ad is typically larger than that of a post-roll ad. Thus, an ad positioning algorithm would have to carefully consider this tradeoff when deciding where to place ads. Our work provides an important input to such an algorithm, though designing optimal ad placement algorithms is beyond the scope of our work. However, our results do show that post-roll ads are generally inferior to mid-roll and pre-roll ads, since post roll ads have both smaller audience sizes and lesser ad completion rates.

5.1.3 Ad Length

We classify each ad into the three common categories, 15-second, 20-second, and 30-second ads, and compute the completion rate for each category (cf. Figure 7). Ads of 30 seconds in length had the highest completion rate and 20-second ads have the least. A fundamental question is how the ad length causally influences its completion rate. With a purely correlational analysis such as that shown in Figure 7, one is liable to incorrectly conclude that 20-second ads are detrimental to ad completion and the sweet spots are 15-second and 30-second ads. Further, the results appear to contradict the intuition that longer ads are more likely to be abandoned, since viewers are more likely to loose patience.

To dig deeper, we analyzed the ad positions of the different ad lengths (cf. Figure 8). We noticed that 30-second ads are placed most often as mid-rolls since advertisers intuitively realize what we quantified in Section 5.1.2 that the viewer is more engaged in the middle of video and tend to place their longest ads there. Thus, the observed high completion rate for 30-second ads could be an influence of its ad position that counteracts its larger length. Further, 15-second ads are placed most often as pre-roll and 20-second ads have a greater chance of being a post-roll than other ad lengths. We have to compensate for the confounding effect of variables such as ad position to isolate the true impact of ad length on ad completion rates. To that end, we design the quasi-experiment described below.

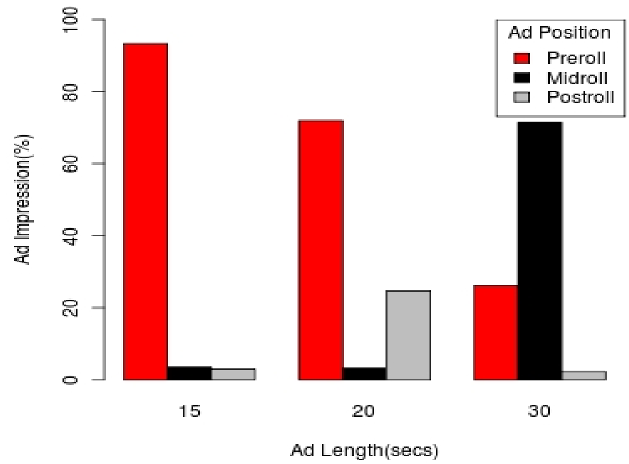

Figure 8: 30-second ads are most commonly mid-rolls, and 15-second ads most commonly pre-rolls. 20-second ads are more often post-rolls than other lengths.

Assessing Causal Impact.

QED. We design a quasi-experiment where the independent variable is the ad length (15-second, 20-second, or 30-second) and the other potential confounding variables are matched. For a given pair of ad-lengths $x \neq y$ and $x, y \in \{15, 20, 30\}$ seconds, we design a quasi-experiment with the treated set consisting of videos that contained an ad of length x and the untreated set consisting of videos that contained an ad of length y. The matching algorithm that we use is similar to that in Figure 6 with the following differences. When forming the matched pair of views $(u, v) \in M$, we ensure that view u played an ad of length x and view v played an ad of length y. To account for the influence of ad position, we ensure that the ads were played in the same position. Further, we ensure that the viewers of u and v are similar with the same geography and connection type and are watching exactly same video. The scoring step is identical to the matching algorithm of Figure 6.

QED Results. The results of the quasi-experiments are shown in Table 6. Our results show that 15-second ads completed 2.86% more often than the 20-second ones in a head-to-head comparison that accounts for the confounding factors. Likewise, 20-second ads completed 3.89% more often than the 30-second ones in the head-to-head compari-

Treated/Untreated	Net Outcome
15 sec/20 sec	2.86%
20 sec/30 sec	3.89%

Table 6: Net QED outcomes support the assertion that longer ads result in fewer completions.

son. Further, using the sign test, the p-value for the quasi-experiment is at most 8.52×10^{-30}, confirming the statistical significance of the results. Thus we state the following rule.

RULE 5.2. *On average, a shorter ad is more likely to complete than a longer ad, when the other confounding factors are neutralized.*

5.2 Impact of Video-related Factors

We examine two factors that relate to the video: its content as identified by its unique url, and its length.

5.2.1 Video Content

People typically watch ads so that they are allowed to watch the video. Therefore, it is reasonable to ask what influence the video itself has on the completion rate of the ads embedded within it. Videos in our traces are uniquely identified by their urls. A video could have been viewed multiple times, and multiple ads could have been shown as part of each view. The ad completion rate of a video is simply the percentage of all ad impressions shown with that video that completed. *Ad completion rate of a video is not to be confused with the unrelated metric of video completion rate that relates to whether the video itself completed or not.* One could imagine that the ad completion rates vary from video to video, with videos with compelling content having high ad completion rates and videos with boring content having lower ad completion rates. In Figure 9, we do indeed see a large variation in ad completion rate across different videos with half the ad impressions coming from videos that have an ad completion rates of 90% or smaller.

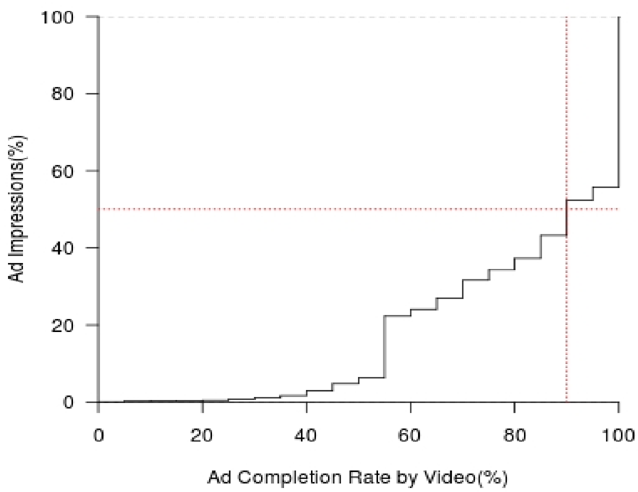

Figure 9: The percentage of ad impressions $y\%$ from videos with ad completion rate at most $x\%$, plotted in 5% buckets of ad completion rate. Half the ad impressions belonged to videos with completion rate 90% or smaller.

5.2.2 Video Length

We narrow our focus to the length of video to assess how it relates to the ad completion rate. We bucket the video length into one minute buckets and compute the average ad completion rate of the videos in each bucket. (Each video is weighted by the number of ad impressions shown with that video for computing the average.) We plot ad completion rate as a function of the video length in Figure 10. The ad completion rate shows a positive correlation with video length with Kendall correlation of 0.23. One can further bucket the videos according to whether they are short-form or long-form and one can see that short-form video has a smaller ad completion rate than the long-form (cf. Figure 11).

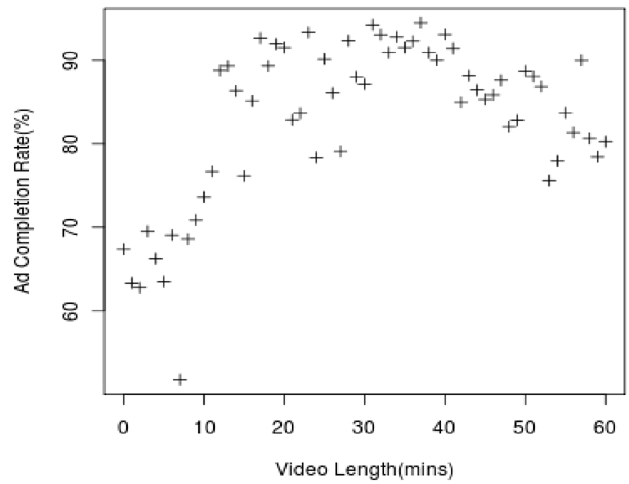

Figure 10: Ad completion rate and video length have a positive correlation with a Kendall coefficient of 0.23.

Our initial correlational results support the intuition that a viewer exhibits more patience for an ad to complete if they are watching long-form content such as a TV episode or a movie that are often perceived to be of greater value than short-form content. Such a phenomena is known to hold in the physical world where researchers who study the psychology of queuing [15] have shown that people have more patience for waiting in longer queues if the perceived value of the service that they are waiting for is greater. Duration of the service often influences its perceived value with longer durations often perceived as having greater value. In [14], we showed that viewers are more likely to wait without abandoning for a longer video to startup than a shorter one. Our current work implies that a similar phenomenon holds for viewer patience for ads to complete.

Assessing Causal Impact.

Based on our correlational evidence above, we would like to establish the following causal rule by a carefully designed quasi-experiment.

RULE 5.3. *On average, placing an ad in long-form video can cause a greater completion rate in comparison to placing the same ad in a short-form video.*

QED. We conduct a quasi-experiment where the independent variable is the video length (long-form versus short-form) and other potential confounding variables are matched.

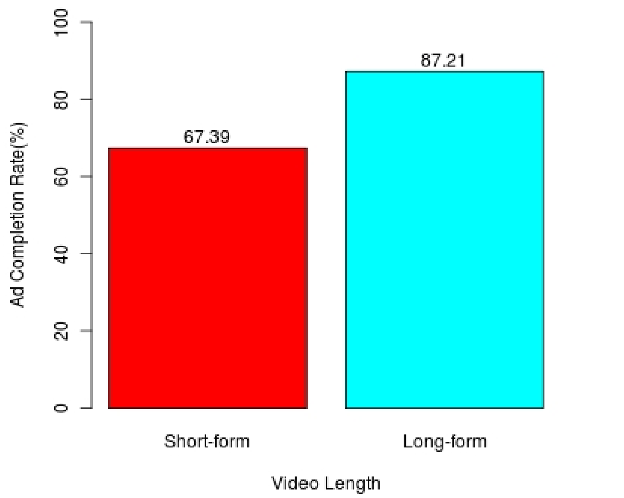

Figure 11: Ads embedded in long-form video such as a TV episode or a movie complete more often than ads embedded in short-form video such as a news clip.

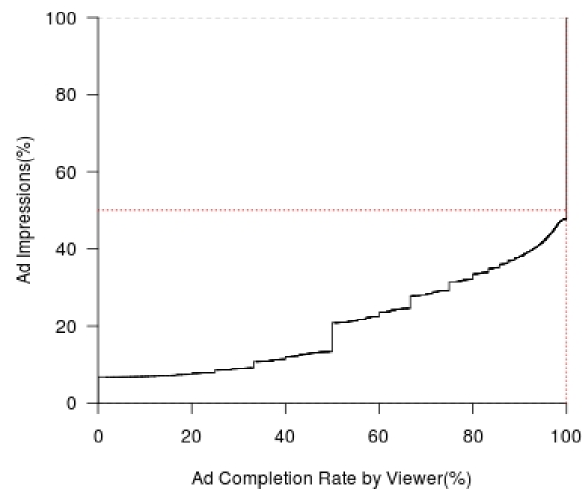

Figure 12: The percentage of ad impressions $y\%$ from viewers with completion rate at most $x\%$.

The matching algorithm that we use is similar to that in Figure 6 with the following differences. Since the independent variable is video length, the treated set consists of long-form videos with ads, while the untreated set consists of short-form videos with ads. When forming the matched pair of views $(u, v) \in M$, we ensure that the paired views played the same ad in the same position, i.e., the ad was pre-roll, mid-roll, or post-roll in both views. Further, the viewers of u and v are similar in that they are from the same geography and have the same connection type. Finally, even though u and v are watching different videos, one long-form and the other short-form, we ensure that they are watching videos from the same video provider. The scoring step is identical to the matching algorithm of Figure 6.

QED Results. The results of the quasi-experiment produced a net outcome of 4.2%. The positive net outcome supports Rule 5.3 by showing that on average an ad that is placed in long-form video is 4.2% more likely to complete than the same ad placed in short-form video. Further, using the sign test, the p-value for the quasi-experiment was at most 9.9×10^{-324}, confirming the statistical significance of the results.

Discussion.

The impact of video length on ad completion rate is confounded by factors such as ad position. For instance, mid-roll ads that tend to have higher completion rates are more commonly embedded in long-form video than in short-form video. Thus, the influence of ad position must be neutralized to get a clearer picture of the impact of video length in isolation. Accounting for such confounding factors in the QED analysis shows a smaller (but still significant) impact of video length, though that impact is smaller than what is implied by the simpler analysis of Figure 11.

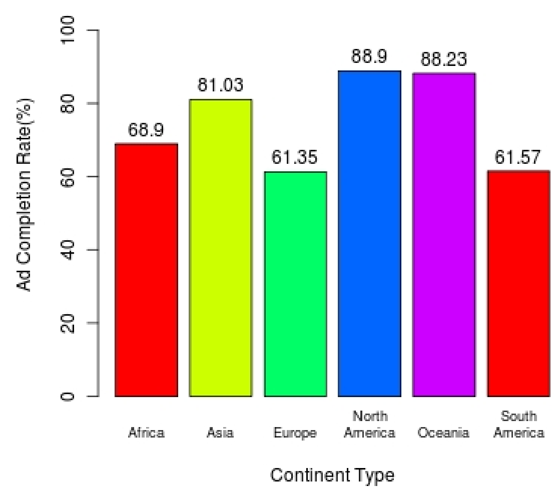

Figure 13: Europe has the smallest completion rate while North America has the greatest.

5.3 Impact of Viewer-related Factors

We examine three factors that relate to the viewer in more depth: the viewer as identified by his/her unique GUID, the viewer's geographical location, and the temporal factors when the ad was played.

5.3.1 *Viewer's Identity*

We compute the ad completion rate of each viewer as simply the percentage of ad impressions that the viewer watched to completion. In Figure 12, we plot the percent of ad impressions $y\%$ that were watched by viewers with completion rate less than or equal to $x\%$. One can notice the concentrations of viewers around completion rates of 0%, 50%, and 100%. These concentrations are an artifact of the fact that a large fraction of viewers see a small number of ads. For instance, 51.2% see one ad contributing to concentrations around 0% and 100%. And, 20.9% see only two ads, contributing to concentrations around integer multiples of 1/2. More generally, one can observe concentrations around integer multiples of $1/i$, where i is a small integer.

5.3.2 *Geography*

In Figure 13, we show the ad completion rates across different continents in the world. Perhaps the most striking contrast are between the two most trafficked continents with Europe having the lowest completion rate and North America having the highest.

5.3.3 *Temporal Factors*

A plausible hypothesis that exists as a folklore is that viewers are more likely to watch ads (and complete them) in the weekend or in the evenings where they tend to be more relaxed, more patient, and have more spare time. Indeed, both video and ad viewership peaks in the late evening as shown in Figures 14 and 15 respectively. However, as shown in Figure 16, ad completion rates did not show much time-of-day variation and were nearly identical between weekday and weekend.

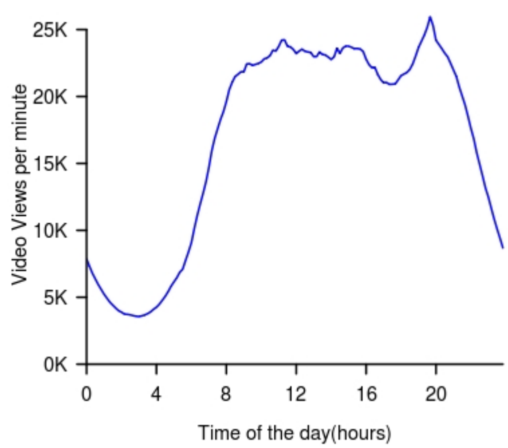

Figure 14: Video viewership is high during the day, dips slightly in the evening, and peaks in the late evening.

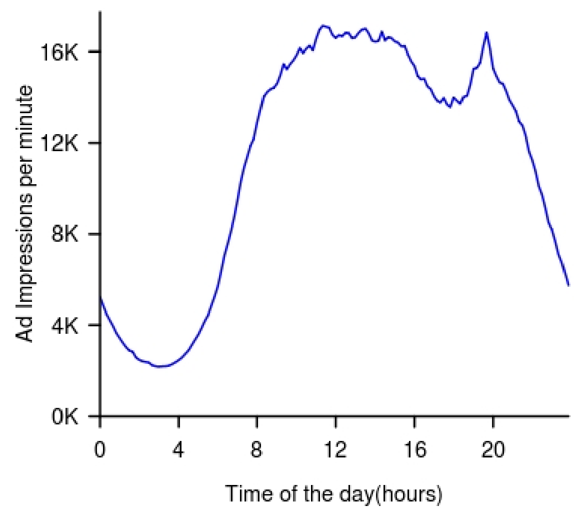

Figure 15: Ad viewership roughly follows the same trend as video viewership.

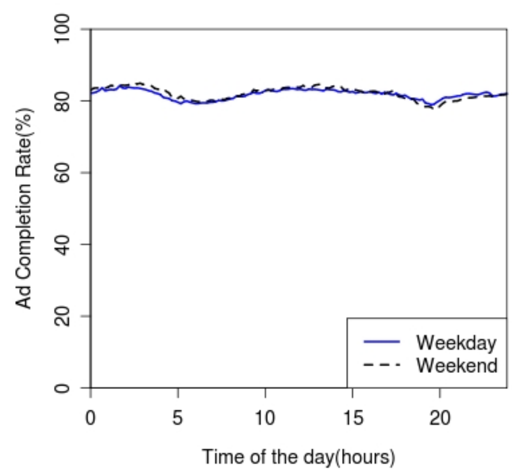

Figure 16: However, ad completion rates do not show major weekday/weekend or time-of-day variations.

5.3.4 *Repeat Visitors versus One-time Visitors*

It is likely that a viewer who repeatedly comes to a video provider's site is more interested in the video content offered by that site than a viewer who comes only occasionally. Would the increased viewer interest translate to a higher ad completion rate? To quantitatively answer that question, for each site we classified each viewer of that site into two categories. A *repeat visitor* is a viewer who made two or more visits to that site during the 15-day period of measurement. A *one-time visitor* made exactly one visit[8] to the site in the 15-day measurement period. Recall that a visit is single session of a viewer visiting a video provider's site to watch videos (and ads). Two consecutive visits to a site from the same viewer are separated by at least 30 minutes of inactivity (cf. Figure 1). As shown in Figure 17, there is a noticeable difference in ad completion rates between repeat visitors and one-time visitors and that difference persists in all three ad categories of pre-roll, mid-roll, and post-roll.

[8]Note that if a viewer came back to the site after the 15-day window, our measurements would not show it.

6. AD ABANDONMENT RATE

While ad completion rates measure whether viewers complete watching an ad or not, ad abandonment rates measure what portion of the ad was played before the viewer

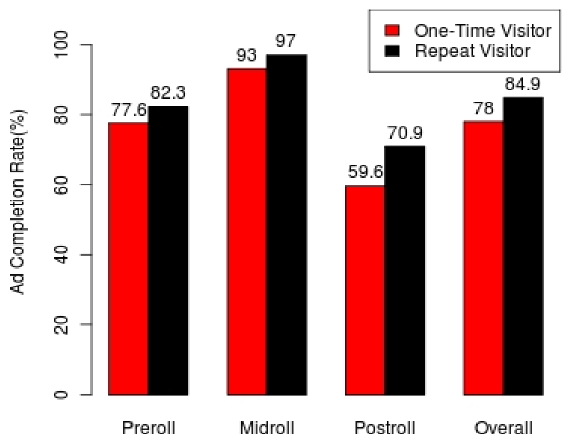

Figure 17: Repeat visitors are more likely to watch ads to completion than one-time visitors to a video provider's site.

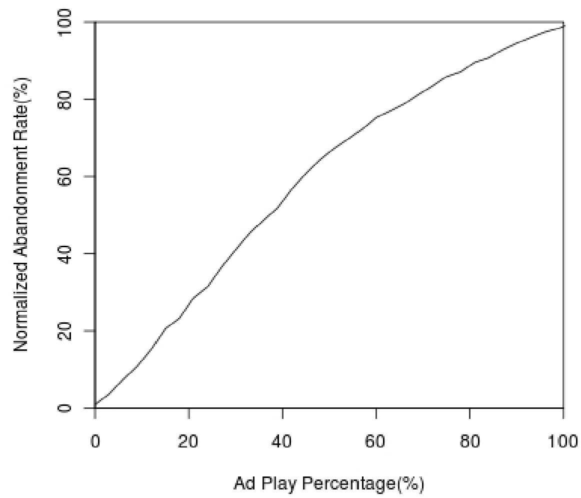

Figure 18: Normalized abandonment rate as a function of ad play percentage has a concave form. Of the viewers who eventually abandon the ad, about a third of the them have abandoned before the quarter-way mark and two-thirds of them have abandoned before the half-way mark.

abandoned. Thus, abandonment rates provide more granular information than completion rates. We define metrics we use to study abandonment. Suppose we have an ad of length L time units. *Ad play time* x, $0 \leq x \leq L$, refers to the amount of time that the ad was played by the viewer during an ad impression. The abandonment rate at time x, $0 \leq x \leq L$, is the percentage of ad impressions that have ad play time less than x, i.e., the percentage of ad impressions where the ad was watched for fewer than x time units. By definition, the abandonment rate of the ad at time $x = L$ is 100 minus that ad's completion rate, since viewers who did not abandon and watched all L time units completed the ad. When aggregating abandonment rates across ads with different lengths, we plot ad abandonment rate as a function of a normalized value called *ad play percentage* which is (ad play time/ad length) $\times$ 100. Further, we define *normalized abandonment rate* to be

$$(\text{ad abandonment rate}/(100 - \text{ad completion rate})) \times 100.$$

Aggregated over all ad impressions in our study, the abandonment rate when ad play percentage equals 100% is 17.9%, which equals 100 minus the system-wide completion rate of 82.1%. In Figure 18, we plot the *normalized* abandonment rate as a function of the ad play percentage. Normalized ad abandonment rate is a concave function with viewers abandoning at a greater rate initially that subsequently tapers off. One can observe from the figure that when 25% of the ad is played, the normalized abandonment rate is already 33.3%, i.e., one-third of the viewers who eventually abandon have abandoned on or before the quarter-way mark in the ad. Likewise, at 50%, the normalized abandonment rate is 67%, i.e., two-thirds of the viewers who eventually abandon have abandoned on or before the half-way mark in the ad.

Next, in Figure 19, we plot the normalized abandonment rate as a function of ad play time to examine how viewers abandon for each of the three ad lengths. By definition, the three abandonment curves reach the normalized abandonment rate of 100% at 15, 20, and 30 seconds respectively. However, the normalized abandonment rates are nearly identical for the first few seconds and diverge beyond that point. This suggests that perhaps a significant fraction of viewers abandon as soon as the ad starts independent of its length.

Finally, in Figure 20, we show the normalized abandonment rate for the different connection types. Our results do not show major differences between the four major connection types for when viewers who eventually abandon stop watching the ad. One *plausible* explanation could be that viewers have a similar expectations on how long they would have to wait for an ad to complete, independent of their connectivity. This could be contrasted with the situation where viewers are waiting for a video to start up after a play is initiated. In this situation, viewers with high-speed connectivity (say, fiber) rightfully expect the video to start up sooner than viewers on a mobile connection. Indeed, in this situation we showed in our prior work [14] that viewers with faster connectivity abandoned the video sooner than those with slower connectivity, presumably because the former had greater expectations for a quicker startup and hence showed less patience for the video to start up.

6.1 Do viewers have more patience for ads than slow-loading videos?

Consider two situations where a viewer must wait to watch the video of his/her choice. Suppose the first set of viewers must wait for a video that is slow to load and start playing due to performance issues. While the second set of viewers must wait for a pre-roll ad to complete before the video begins. We compare the rate at which viewers lose patience and abandon in both situations. Figure 21 shows the aban-

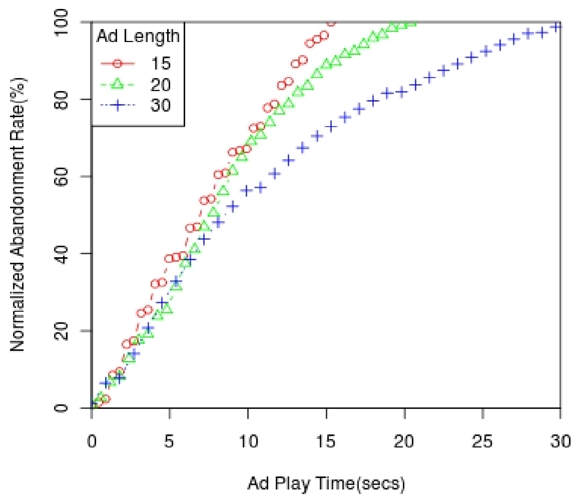

Figure 19: Normalized abandonment rate for different ad lengths.

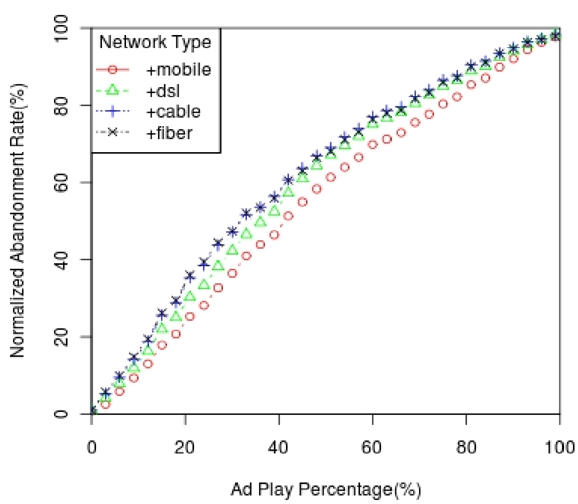

Figure 20: Normalized abandonment rates are roughly similar for the different connection types.

donment rate[9] for slow-loading videos from our earlier work in [14] in comparison with the abandonment rate seen for pre-roll ads from our current study. In both cases, as viewers wait more, i.e., delay is higher, more people abandon. However, it is easy to see that viewers abandon much faster when they are waiting for a slow-loading video than when watching a pre-roll ad. For instance, even though the waiting time is the same, at the 10-second mark, viewers from the first set have abandoned at a rate more than three times higher than that of the second set.

[9]Note that we use the absolute value of the abandonment rate, rather than the normalized one.

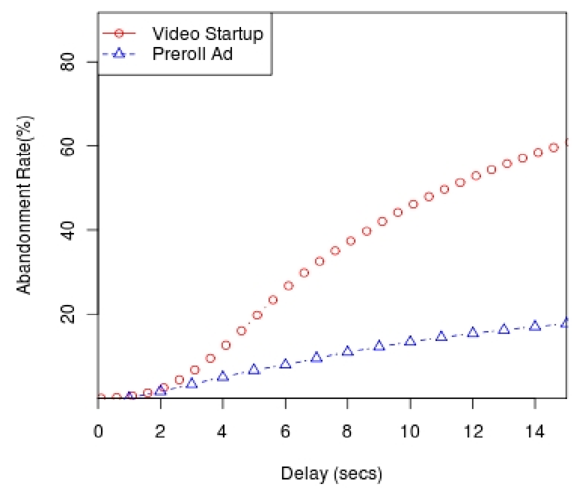

Figure 21: Viewers abandon at a rate more than three times greater for a slow-loading video than for a pre-roll ad. At the 10-second mark, 45.8% of the viewers waiting for the slow loading video had abandoned, compared to only 13.4% of the viewers who abandoned the pre-roll ad.

We conjecture that a reason for the drastically different abandonment rates is viewer psychology. In the physical world where people wait for a service, it is well-known that people are more frustrated with waits that are unexpected or are of uncertain duration. Our results demonstrate that the same is true in the online world. Viewers opt in to watch a pre-roll ad and often view it as an implicit form of payment for the content. Thus, the viewer *expects* to wait for the ad to complete and the wait is often of a predictable duration.Thus, the frustration and the resulting abandonment is smaller. Whereas having to wait for a slow-loading video to start playing is often unexpected and the wait itself is of unknown duration. Thus, the frustration and the resulting abandonment is greater. *This result quantitatively suggests that if the media player can predict that a video will start up late due to an underlying network problem, one can likely reduce viewer abandonment by inserting an ad of the appropriate length instead of showing a blank screen or a "spinning wheel".*

7. RELATED WORK

We are not aware of large-scale scientific studies of video ads and their impact akin to our work. However, given its importance, the metrics that we study such as ad completion rate, abandonment rate are widely reported on a quarterly or yearly basis by ad networks such as FreeWheel [5], Adobe[1], and Bright Roll [3] and analytics providers such as comScore [4]. Since the business of online video relies on ad completion rates, audience size and other such metrics, the major industry standards body IAB [6] provides guidelines on how such video monetization metrics ought to be measured. Our work on systematically understanding the impact of various factors on ad viewing behavior and extracting general rules via quasi-experiments is unique and significantly contributes to our scientific understanding of ad efficacy and video monetization. There has recently been research on understanding the impact of video performance on viewer behavior [11, 14], and in the use of client-side measurements for better video delivery [17]. These works share a commonality with our current work in the sense of using large amounts of data collected from media players, but are targeted towards very different research goals.

In terms of the techniques, our prior work [14] used quasi-experiments in a network measurement setting. In this paper, we develop the QED technique further and use it in a different context for the study of video ads. While seldom used in measurement studies of networked systems prior to our work in [14], quasi-experiments have a long and distinguished history of use in the social and medical sciences that is well documented in [20].

8. CONCLUSIONS

To our knowledge, our work is the first in-depth scientific study of video ads and their effectiveness. We explored how ad effectiveness as measured by ad completion rate is impacted by key properties of the ad, of the video, and of the viewer. A key contribution of our work is that we go beyond simple characterization to derive causal rules of viewer behavior using quasi-experimental designs (QEDs). We show that an ad is 18.1% more likely to complete when placed as a mid-roll than as a pre-roll, and 14.3% more likely to complete when placed as pre-roll than as a post-roll. Next, we show that completion rate of an ad decreases with increasing ad length. A 15-second ad is 2.9% more likely to complete than a 20-second ad, which in turn is 3.9% more likely to complete than a 30-second ad. Further, we show that the ad completion rate is influenced by the video in which the ad is placed. An ad placed in long-form videos such as movies and TV episodes is 4.2% more likely to complete than the same ad placed in short-form video such as news clips. We also studied the abandonment rate metric and showed that viewers abandon more quickly in the beginning of the ad and abandon at slower rates as the ad progresses. Our work represents a first step towards scientifically understanding video ads and viewer behavior. Such understanding is crucial for the long-term viability of online videos and the future evolution of the Internet ecosystem.

9. ACKNOWLEDGEMENTS

We thank Girish Bettadpur for his deep insights into the video ecosystem, Rick Weber for providing computing resources, and Harish Kammanahalli for his support. Any opinions expressed in this work are solely those of the authors and not necessarily those of Akamai Technologies.

10. REFERENCES

[1] Adobe Auditude. http://goo.gl/iI7Mc.

[2] *Akamai Media Analytics*. http://goo.gl/lnFwF.

[3] Bright roll. http://www.brightroll.com.

[4] comscore. http://www.comscore.com.

[5] Freewheel. http://www.freewheel.tv.

[6] Interactive Advertising Bureau (IAB). http://www.iab.net.

[7] Video Plaza. http://www.videoplaza.com.

[8] Cynthia Boris. What is the fastest growing online ad format? Video! http://goo.gl/ylnX8.

[9] Cisco. Visual Networking Index. http://bit.ly/KXDUaX.

[10] John Dilley, Bruce M. Maggs, Jay Parikh, Harald Prokop, Ramesh K. Sitaraman, and William E. Weihl. Globally distributed content delivery. *IEEE Internet Computing*, 6(5):50–58, 2002.

[11] Florin Dobrian, Vyas Sekar, Asad Awan, Ion Stoica, Dilip Joseph, Aditya Ganjam, Jibin Zhan, and Hui Zhang. Understanding the impact of video quality on user engagement. In *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 362–373, 2011.

[12] eMarketer. Click rates complicate online video ad metrics. http://goo.gl/vPGzs.

[13] John T Kent. Information gain and a general measure of correlation. *Biometrika*, 70(1):163–173, 1983.

[14] S Shunmuga Krishnan and Ramesh K Sitaraman. Video stream quality impacts viewer behavior: inferring causality using quasi-experimental designs. In *Proceedings of the ACM conference on Internet measurement conference*, pages 211–224. ACM, 2012.

[15] R.C. Larson. Perspectives on queues: Social justice and the psychology of queueing. *Operations Research*, pages 895–905, 1987.

[16] E.L. Lehmann and J.P. Romano. *Testing statistical hypotheses*. Springer Verlag, 2005.

[17] X. Liu, F. Dobrian, H. Milner, J. Jiang, V. Sekar, I. Stoica, and H. Zhang. A case for a coordinated internet video control plane. In *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 359–370, 2012.

[18] E. Nygren, R.K. Sitaraman, and J. Sun. The Akamai Network: A platform for high-performance Internet applications. *ACM SIGOPS Operating Systems Review*, 44(3):2–19, 2010.

[19] H. Oktay, B.J. Taylor, and D.D. Jensen. Causal discovery in social media using quasi-experimental designs. In *Proceedings of the First Workshop on Social Media Analytics*, pages 1–9. ACM, 2010.

[20] W.R. Shadish, T.D. Cook, and D.T. Campbell. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company, 2002.

[21] D.A. Wolfe and M. Hollander. Nonparametric statistical methods. *Nonparametric statistical methods*, 1973.