Amped about AMP? Time to look more carefully

Authors: Bradley B Doll, Chris Breaux, Kris Harbold

The Daily Beast
Chartbeat

Correspondence should be addressed to: brad.doll@thedailybeast.com and christopher@chartbeat.com

Abstract
Numerous reports suggest that the Accelerated Mobile Pages (AMP) project can increase publisher traffic because of special promotion by Google. We present the first formal statistical report on the relationship between AMP adoption and publisher traffic. We found that AMP adoption resulted in a 22% average Google traffic increase for the 159 publishers we analyzed. But, because website traffic is very noisy, and the impact of AMP is relatively small, only 1 in 3 publishers were able to see an unequivocal positive traffic impact from moving to AMP. Our findings suggest that publishers seeing poorer monetization of AMP relative to standard web pages will struggle to make the case that a traffic boost from AMP makes up for lost revenue.

Introduction
The Accelerated Mobile Pages project (AMP), championed by Google, has been extensively covered in the press since launch. While the format’s fast page load times have been celebrated (1), response to its monetization have been more mixed (2, 3, 4). Achieving parity between AMP and standard web page revenue has been a challenge for some publishers. This difference may be made up by extra traffic from the increased promotion Google offers AMP publishers, and a number of observations suggest AMP indeed increases Google referrals (5, 6, 7). However, these reports lacked important statistical controls, leaving open the question of whether and by how much AMP affects publisher traffic, and therefore whether and how it affects revenue. We resolved this issue via formal statistical analysis of traffic to The Daily Beast and 158 other publishers who launched AMP in 2017. We found evidence that, on average, AMP increases the mobile traffic a publisher receives from Google by 22% (+/-4%). However, we found substantial variability across publishers such that 66% (including The Daily Beast) saw no clear evidence of an increase. Moreover, these effects varied widely, with individual publishers seeing effects ranging from +151% to -58%. These findings, together with potentially lower revenue from AMP relative to standard web pages, suggest that publishers take a careful look at AMP returns when deciding on adoption.
Results

World events drive large and typically unpredictable shifts in publisher web traffic, making it difficult to understand whether apparent changes in visits are caused by publisher efforts or news volatility. To separate effects of Google AMP on traffic from other noise, we employed two approaches. The first approach is a controlled experiment conducted at The Daily Beast in which content was randomly assigned to be AMP eligible and traffic was compared between eligible and ineligible content. This random assignment controls for the confounding effect of news events because these unpredictable events are equally likely to fall into either experimental group. The second approach is an analysis on traffic time series for the Daily Beast and 158 anonymous Chartbeat customers. This analysis assesses the change in each site’s Google traffic after AMP launch and compares it to the change in traffic from other referral sources. Comparison of Google to other sources controls for unpredictable news events because they drive similar traffic shifts for all traffic sources (Figure 1). Evidence of AMP uniquely affecting traffic requires that changes in Google sourced sessions exceed the changes in sessions from a site’s other referrers.

Analysis 1: The Daily Beast A/B test

To understand the impact of AMP on traffic, The Daily Beast ran an A/B test from 1-24-2018 to 3-5-2018 in which half of newly published articles were randomly made AMP eligible. In this time window 582 AMP and 566 non-AMP articles were published. Compared to launching AMP and comparing traffic before and after (Figure 1), this test offers greater statistical power in a shorter period of time with fewer possible confounds. Because AMP may drive incremental sessions via Top Stories news carousels and other promotion, we counted Google-sourced mobile device page views for AMP or non-AMP content only when the content was the first page view in a session. This removes confounding effects of onsite recirculation on page views.

We adopted a Bayesian approach to model the effect of AMP, and uncertainty about it (see Methods). This approach provides a distribution of plausible effect sizes conditional on a specified model and observed data. This distribution of effects for the average AMP article had a positive mean (+10%), but with high uncertainty (Figure 2) suggesting AMP may not boost traffic, and in fact may even reduce it. Specifically, the 95% credible interval on the estimated AMP effect ranged from -8.4% to +31.7%. In total, 15% of the plausible effect sizes for AMP were less than or equal to 0% change, and the widest central interval of effect sizes excluding zero was 66%, providing weak evidence at best for a positive effect of AMP on traffic.

For a publisher, the key question about AMP is likely not how it changes traffic, but rather how it changes revenue. These questions have different answers because of potentially lower ad revenue on AMP compared to standard mobile web pages (2, 3, 4). The Daily Beast calculated that a traffic increase from AMP of 23% would be required to make up for the lost revenue. Our results show that an increase in traffic this large is very unlikely, with only the top 11% of plausible AMP effect sizes reaching or exceeding this threshold (Figure 2).

Analysis 2: Time series analysis of The Daily Beast and 158 Chartbeat customer sites
While clear affirmative evidence of an AMP traffic effect is lacking for The Daily Beast, publishers may likely vary (2, 8). We sought to assess whether AMP effects on traffic differed across publishers by analyzing the traffic for 158 publishers working with Chartbeat who launched AMP in 2017 together with The Daily Beast’s initial AMP launch. This cohort of publishers consists of a mix of local and national publishers of news, sports, and lifestyle content. A clear majority of these publishers are US-based, but 10 countries are represented. In contrast to the A/B test, comparing site traffic from different time periods presents special challenges (see Methods for full details of statistical controls). Website traffic changes are often abrupt and unexpected, such that a traffic increase after AMP launch may have nothing to do with AMP, but rather with an increase in news events or seasonal changes that draw traffic. We reasoned that such events tend to drive traffic changes similarly across all sources (Figure 1). Therefore, if AMP increases traffic, we would expect a greater increase in Google-sourced mobile sessions than in sessions from non-Google sources after AMP launch.

In this analysis, we found that only 34% of individual publishers saw clear evidence that AMP positively affected traffic (i.e. their 95% confidence intervals were greater than zero. Figure 3). Though the average effect of AMP was +22% (95% credible interval = 22% +4-4%), individual publisher estimates varied substantially from the average +151% to -58% (Figure 3). In order to better identify the true AMP effects within and across publishers, we incorporated the uncertainty of each publisher’s AMP effect into the average calculation (see Methods). This analysis can adjust the degree to which publisher effects impact the average AMP effect proportional to their uncertainties, and thus, the average itself. The average effect was essentially unchanged 21% (+-4%), but uncertainty about individual publisher effects was reduced, with 42% of publishers showing clear positive effects of AMP (95% credible intervals greater than zero). While these latter results provide more optimism about AMP, they are challenging for individual publishers to apply because they rely on multi-publisher data. Further, publishers still differed widely from this average, ranging from +81% (+-27%) to -42% (+-13%; Figure 4).

In contrast to the A/B analysis, these multi-publisher analyses are correlational, leaving open greater possibility that some unknown factor caused a relative increase in Google traffic, which we erroneously identified as an AMP effect here. Of particular concern is the fact that the majority of publishers in our data set launched AMP in late summer 2017 (72% in August and September), coinciding with a previously observed increase in Google traffic among Chartbeat customers (5). Thus, it is possible that AMP has no true effect, but we observed one by analyzing sites that happened to launch AMP when Google referral traffic increased. We reasoned that if the +21% average effect we identified was in fact caused by AMP, we should see no such effect for sites that never launched AMP. In our control analysis, we took data from 158 Chartbeat customers who have not launched AMP, randomly assigned them to the AMP launch dates for the sites in our main analysis, and analyzed the traffic time series for each publisher (see Methods for details). We averaged the AMP effect estimates across publishers, and repeated the random assignment, analysis, and averaging process 10,000 times. The
resulting distribution of effects sets our expectation for the size of the Google traffic shift in absence of any AMP effects. This distribution had a mean of -5% and its central 95% ranged from -36% to +23%. The +22% effect observed for the AMP group is relatively extreme in this distribution (95th percentile), as would be expected if the Google traffic boost was due to AMP, rather than from an unrelated increase.

Our control analysis featured non-AMP sites that were similar to the AMP sites we analyzed (barring the adoption of Accelerated Mobile Pages). However, it remains possible that these groups differed in some unmeasured way that explains the difference in Google traffic changes observed in the two cohorts. To more fully assess our confidence in the AMP cohort result, we separated the non-Google comparison traffic into three large groups: Facebook, Direct, and Other (the latter capturing all remaining referral sources), and repeated our analysis comparing Google against each of these groups. We looked for evidence that Google traffic increased after AMP launch relative to each of these different sources as supporting the presence of a true effect. In this analysis, Google showed the largest increase compared to Facebook (+25% increase +4%, 40% of publishers with clear positive effects, publisher mean effects ranging from -30% to +89%). This result is consistent with publisher reports of declining Facebook traffic (9, 10, 11). However, the main effect of AMP was not driven entirely by a Facebook decline, as we also observed positive average effects of Google compared to Direct (18% increase +4%, 28% of publishers with clear positive effects, publisher mean effects ranging from -12% to +78%) and Google compared to Other (9% increase +5%, 25% of publishers with clear positive effects, publisher mean effects ranging from -59% to 193%). These results further confirm the presence of a real positive effect of AMP on traffic, but one that differs substantially across publishers (Figure 5).

Finally, we sought to understand the source of variability in the AMP effect across publishers by correlating the measured change in Google traffic with a numerous candidate explanatory factors. However, we found no statistically significant correlation between AMP effect sizes and publisher features we possess: 2017 implementation date (r = 0.10, p = 0.22), overall traffic to site (r = -0.10, p = 0.23), share of search traffic (r = -0.14, p = 0.08), or broad type of content (one-way ANOVA: p = 0.44, η² = 0.01).

Discussion
Numerous reports have estimated the impact of AMP on publisher traffic (5, 6, 7), but these analyses lacked important statistical controls and contained no measures of uncertainty. Thus, it has remained unclear whether AMP truly has a positive impact on traffic, and if so, what size of boost should publishers expect. Here, through formal statistical analyses, we estimated an average AMP effect size of 21% (+4%) with individual publishers varying widely around this estimate, and only a minority showing clear evidence of an increase. The Daily Beast was in the majority group, and also showed no clear AMP effect in a substantially more powerful A/B test. Though we assessed a number of possibilities, the source of the variability across publishers remains unexplained.
Google’s promotion of AMP continues, with search results carousels proliferating (12), AMP content serving in Google’s re-launched newsstand (13), and new products appearing such as AMP stories (14) and AMP email (15). AMP adoption is also being encouraged more subtly. In July, Google started using page speed as a ranking factor for search results (16). AMP is not specifically required by this Speed Update, but it is a clear path to fast pages. Given these developments, publishers will likely need to consider the format to remain competitive. However, our results suggest this may be an unhappy reality for those with AMP monetization challenges, because it is difficult for an individual publisher to understand their AMP traffic change and whether it makes up for lost revenue.

In our data set, we found that only 34% of publishers on their own would be able to make a clear case for an increase. This number improved to 42% when effect and uncertainty estimates across publishers were entered into the same analysis, but individual publishers typically have no provisions to complete this more powerful type of analysis. Publishers might seek to increase statistical power to understand effects by collecting more data. However, data collection times here for both the time series and the more powerful A/B test already require long waits in the context of a business decision (and, with longer collection times, more confounds can appear and plague these tests).

The source of variability in the AMP effect between publishers remains unclear. In our analyses, this variability did not relate to site traffic volume, proportion of search traffic, or publisher content type. One possibility is that AMP promotion asymmetrically benefits publishers, unleveling the playing field, as some have asserted (17). Indeed, Search Engine Land reported that website promotion via the AMP Top Stories news carousel primarily benefits the first result card. Clicks to subsequent carousel cards drop off more quickly than for successive entries on standard search result pages (8). Another possibility is that AMP implementation varies in quality. Stone Temple Consulting reported that sites showing better returns from the technology had invested more time in making AMP and non-AMP pages similar (18). While we do not possess implementation details on the other sites in our analysis, The Daily Beast executed the recommended high quality implementation and saw no clear evidence of a traffic increase.

Because we lack page monetization data for the Chartbeat customers in our data set, our analyses focused on whether and by how much AMP increases a publisher’s Google traffic. From a publishers point of view, however, the key question of interest is likely whether AMP is a revenue neutral or better business decision. Several publishers have reported reduced revenue on Accelerated Mobile Pages (up to 50%), while others are seeing similar returns to standard web pages (2). For publishers in the former group, lost revenue can be accounted for by a traffic boost, or potentially by optimization of ads on AMP. The Daily Beast, for example, requires a 23% increase in traffic to break even on revenue, but 89% of the plausible AMP effect sizes for the site were below this value. The chance of breaking on AMP with a traffic boost is smaller than the chance that AMP boosts traffic at all for The Daily Beast.
In sum, the prospect of AMP is unlikely an positive one for many publishers, with few seeing a clear boost in traffic and some facing substantial monetization challenges (2, 3, 4). Though the technology offers rightly lauded page speed improvements, and potential opportunities in new products, these may come at a high cost for some. As AMP is under continued development, and Google has shown interest in helping publisher efforts (19), discrepancies between AMP and standard web page monetization may be eliminated (20), permitting wider adoption of the technology with reduced revenue risk.

Methods

Analysis 1: The Daily Beast A/B test

The A/B experiment was conducted at The Daily Beast from 1-24-2018 to 3-5-2018. During this time, half of content published to the “article” template was randomly made AMP eligible. No other page types were eligible for the AMP format. For these 1148 articles (582 AMP eligible), we compared the difference in mobile, Google-sourced page views to AMP eligible and AMP ineligible content. We present results for device=mobile and source=google, but we found highly similar results if page views from tablet, and all Google sources (e.g. Google news) were included. To remove effects of onsite recirculation, we only counted a content page view if it was the first in a session (i.e. a “landing” page). Note that recirculation may likely differ between AMP and non-AMP, as the former removes The Daily Beast’s progressive scroll and adds side swipes which take users across publishers. AMP may plausibly shorten sessions compared to mobile web. However, we do not have “stitched” session data for AMP sessions (21), so the current study cannot quantify this effect.

We fit a Gamma linear regression model to the page view count data using Markov Chain Monte Carlo (MCMC, specifically, Hamiltonian Monte Carlo implemented in the Stan probabilistic programming language (22), via the Rstan package (23)). The model has an intercept capturing the average page views to an AMP-ineligible article, and a slope capturing the difference in AMP eligible and ineligible page views. The gamma model was parameterized with shape and rate parameters, with the log link linear model folded into the latter as shape / exp(intercept + slope * AMP eligible indicator variable). We chose as minimally informative priors: intercept ~ normal(0, 10), slope ~ normal(0, 1), and gamma distribution shape parameter shape ~ uniform(0.001, 100).

For this model, and the MCMC models described below, we collected 10,500 samples from the posterior distributions of parameter values (3 chains of 4000, discarding the initial 500 warmup samples in each). We verified chain convergence for each parameter by visual inspection and by assessing the “potential scale reduction factor”, Rhat (23; all Rhat values = 1 in all models, consistent with convergence).
We took quantiles of the posterior parameter distributions as credible intervals. By analogy to the standard 95% confidence interval widely employed in frequentist statistics, we identified the region from the 2.5th to the 97.5th percentile and assessed the values they ranged over, and particularly whether they contained zero. While the exact size of this interval is arbitrary in Bayesian and frequentist statistics, 95% is a ubiquitous indicator of sufficiently strong evidence to pass peer review in many scientific disciplines, and the likely threshold for individual publishers conducting their own statistical analyses of AMP effects, so we report it here for all analyses. To provide the richer, more nuanced picture that accompanies Bayesian inference, we report additional details on the shape of the posterior distributions of AMP estimates and how key variables relate to them.

**Analysis 2: Time series analysis of The Daily Beast and 158 Chartbeat customer sites**

The data set for the time series analysis consisted of mobile session counts by source for The Daily Beast and 158 anonymous Chartbeat customers who launched AMP in 2017. For each publisher the data included the 2 months before and after AMP launch.

The publishers included in this study span from thousands to hundred millions of monthly pageviews, with the median publisher represented having hundreds of thousands of pageviews each month, with 15% of these referred by Google. Notably, this sample does not constitute a random sample of AMP-enabled publishers on the Chartbeat network; many of the largest Chartbeat customers using AMP implemented in 2016. The customers range from local news publishers to sports sites to lifestyle and special interest, and are based in 10 different countries, with the majority coming from the US. The AMP implementation dates range from March 2017 to December 2017, with a notable concentration (72%) in August and September 2017.

Because session information is not tracked by default for AMP page views, session counts can be falsely inflated. Specifically, recirculation from an AMP page gets counted as a new session with source, ampproject.org. To avoid this issue, we examined the effect of removing these sessions with this source from the data set. The Daily Beast conducted the analysis both ways and found negligible differences, perhaps because AMP has no affordance for progressive scroll, a core recirculation mechanism for the site. The Chartbeat results include these sessions (in the non-Google group, and in the “Other” group), which should lead to the AMP effect estimates being understated (through analyses comparing Google to Facebook and Direct are unaffected).

Another noteworthy data limitation is that sessions originating from homepages and section fronts are not included in the Chartbeat customer data. Depending on traffic changes to these page types during the time frame in which the data were analyzed, this could cause over or under estimation of the AMP effect (which could bias the analyses comparing Google to non-Google, and Google to Direct).
For each publisher’s daily mobile session counts we fit a Gamma regression model with log link using maximum likelihood estimation via the Python package Statsmodels (we chose this approach rather than MCMC to ease computational load). Each model contained independent variables to address nuisance structure in the time series to guard against bias in standard error estimation. Specifically, we included a weekend indicator variable, and 5 variables for traffic on the past 5 days. These specific variables arose through experimentation to best reduce autocorrelation and structure in residuals for modeled traffic to The Daily Beast. We also spot checked residuals on Chartbeat customer sites and found similar results. To assess our question of interest, we included indicator variables for traffic source, and for the time period with live AMP. In order to compare variables across sources within a single model for a publisher, we removed the model intercept and entered each source alone and each source interacted with all the other variables (none of which was entered into the model alone). This nests as many regressions as there are sources into the same model, with each lone source coefficient serving as the intercept for the nested model. This arrangement permits post hoc statistical comparisons of Google and the other sources for each publisher. Specifically, we computed t-tests for the difference in the coefficients of two types of interactions: 1. Interaction of the Google indicator and the live AMP indicator, 2. interaction of the non-Google indicator(s) and live AMP indicator. We fit 2 models to each publisher. The first contained a source indicator for Google and non-Google. The second broke the non-Google data into Direct, Facebook and Other (the latter containing all other referral sources).

We assessed the results of the t-tests by transforming the differences into proportion-change units, and assessing them at the population level. First, we found the percentage of individual publishers with \( P \) values less than 0.05 and entirely positive 95% confidence intervals (“significant” positive frequentist effects). Next, we entered the transformed estimates into a Gaussian model of the average effect. This model contained a parameter for the mean AMP effect, and its standard deviation (estimated via MCMC). We took as weakly informative priors mean AMP effect \( \sim N(0,1) \) and standard deviation \( \sim \text{Cauchy}(0, 2.5) \). We assessed the central 95% credible interval of the posterior parameter samples as described above. This model estimates the AMP effect for publishers on average and the uncertainty of that average.

We additionally assessed the group-level AMP effects by modeling both the individual publisher effects and the uncertainty around them (standard errors transformed into proportion-change units). Here, we modeled the observed AMP effects to arise from a Gaussian distribution with unknown mean and known standard deviation (standard error). This unknown mean was then modeled as a Gaussian distribution with a mean and standard deviation parameter (with priors mean \( \sim N(0,1) \) and standard deviation \( \sim \text{Cauchy}(0, 2.5) \)). This arrangement replaces the publisher level observed data points with distributions in the model of the mean, and allows their spread to affect the group level mean AMP effect estimate, and vice-versa.

The bulk of the 158 Chartbeat publishers in our data set launched AMP in late summer 2017 (72% launched in August and September), at which time, an increase in Google traffic was previously observed (5). To guard against the possibility that our AMP result reflects sampling
error rather than a true AMP effect, we repeated our analysis on a cohort of 158 sites that do not publish Accelerated Mobile Pages. We randomly assigned these publishers to the AMP launch dates for the sites in our main analysis, and fit the model described above to each publisher. We averaged the AMP effect estimates across publishers, and repeated the random assignment, analysis, and averaging process 10,000 times. The resulting distribution of effects sets our expectation for the size of the Google shift absent AMP effects. Placement of the average AMP cohort effect in this distribution indicates how rare it is, and therefore how likely it is to have arisen by random variation rather than by a true AMP effect.

To construct the control group, we took data for more than 1,000 customers who met the following criteria. First, they had never implemented AMP, as determined by lack of web recirculation from an domain.ampproject.org referrer and not having an AMP analytics implementation. Second, the sites had recorded historical data throughout the time periods used in the analysis (that is, January 2017 through February 2018). Third, these sites fell within the range of traffic size in the cohort of AMP sites. From this set of customers, we sampled a set of 158 domains that approximated the traffic size distribution of the AMP cohort.
Figure 1. Changes in traffic tend to affect referrers similarly. Plot depicts the 7-day rolling average of The Daily Beast mobile session count (smoothed to enhance viewability). Because unpredictable shifts in traffic co-occur across sources, their effects on traffic can be controlled for. The Daily Beast launched AMP site-wide on 9-28-17. Google traffic increases to its highest points in the post-launch part of the year. However, the post-launch increase in traffic from other sources (not_google) is similar, suggesting the increase may be related to news events rather than the launch of Google AMP.
Figure 2. Distribution of plausible AMP effects on The Daily Beast estimated from statistical model of A/B test data. Higher bars indicate more plausible effect sizes. Under conventional approaches, the central (densest) 95% of the distribution must exclude zero to suggest a true effect. 95% credible interval here overlaps zero (-8.4%, +31.7%), suggesting no clear effect of AMP on traffic. The widest central credible interval excluding zero is 66%. Grey area indicates the 15% of the distribution at or below zero. Green area indicates the necessary effect sizes for AMP to account for or exceed revenue loss per AMP page with incremental traffic. These breakeven values are even less plausible than a non-zero effect of AMP on traffic.
Figure 3. Estimates of AMP effects for 158 Chartbeat customers and The Daily Beast (in red). Circles represent estimates, vertical lines represent 95% confidence intervals around those estimates. The average of the estimates was 22% (+-4%), but only 34% of individual publisher estimates showed clear evidence of a positive AMP effect (i.e. intervals greater than zero).

Figure 4. Estimates of individual AMP effects for 158 Chartbeat customers and The Daily Beast (in red). Estimates here are conditioned on within and across publisher uncertainty, which reduces extreme observations and adjusts individual publisher uncertainty. Circles represent the average estimated value for each publisher, vertical lines represent the central 95% of the distribution of estimated values (i.e. 95% credible interval). The average AMP effect in this analysis was 21% (+4%), with 42% of publishers showing clear statistical evidence of increased Google traffic after AMP launch. Order from Figure 3 preserved.
Figure 5. Estimates of individual AMP effects for 158 Chartbeat customers and The Daily Beast (in red), with each of the three panels indicating the size of Google increase relative to another traffic referral source: Facebook, Direct, and Other (all remaining referrers). Estimates here, as
in Figure 4, are conditioned on within and across publisher uncertainty, which reduces extreme observations and adjusts individual publisher uncertainty. The average across publishers for each comparison exceeds zero, but a minority of individual publishers show unambiguously positive effects (i.e. 95% credible intervals greater than zero).
References

1. https://medium.com/@chunbuns/google-amp-yields-600-increase-on-mobile-site-page-load-speed-ca8489d815ac
15. https://blog.google/products/g-suite/bringing-power-amp-gmail/
17. www.amplettor.org