

Star Digital: Assessing the Effectiveness of Display Advertising

Brady Engelke, Minya Na, Yidan Gao, William Wu, Chia-Hsuan Chou

2/12/2020

Contents

0.0.1	Executive Summary	1
0.0.2	Exploratory Analysis & Experimental Design Checks	2
0.0.3	Is there a frequency effect of advertising on purchase?	5
0.0.4	Which site should Star Digital advertise on?	8
0.0.5	Is online advertising effective for Star Digital?	8
0.0.6	Final Recommendations	8

0.0.1 Executive Summary

The evolution of the internet and how people interact with it has demanded Star Digital to consider pivoting a substantial portion of their \$100M annual advertising budget towards display advertising. Although online advertising generally facilitates an enhanced ability to measure the ROI of advertising, there are still limitations advertisers must consider. Therefore, a controlled experiment must be conducted for an advertiser to understand what the normal behavior of a consumer would be without seeing ads to quantify the impact of ads on the customer base's purchasing behavior. Star Digital's marketing team did exactly this; the details of the experimental design are delineated below.

The experiment's objective was to increase the subscriptions of packaged video service sales. The independent variables were the number of ad impressions a consumer saw across 6 different websites. Specifying the control group was arguably the most intricate aspect of the design since Star Digital is in business to make money and took care to ensure the experiment was conducted most cost-effectively. The customers that were selected for the control group only saw charity advertisements throughout the campaign while the test group was exposed to the full scope of the campaign. Star Digital spent the same amount of money on the charity ads as the Star Digital ads, forfeiting the potential influence the Star Digital ads could have had on the customers in the control group. Star Digital specified 10% of consumers to be in the control group and 90% to be in the test group. The campaign delivered 170 million impressions to approximately 45 million customers for two months in 2012. The conversion rate for the entire population was 0.153%. A choice-based sample was drawn from this data for experimental purposes that were 50% purchasers, 50% non-purchases while the detail of whether the customer was in control or test was randomized. Star Digital has reached out to the Carlson Consulting Team to understand whether the frequency at which they advertise as well as the specific sites they advertise on impact the effectiveness of the campaign to ultimately determine whether online advertising would be a profitable endeavor.

The Carlson Consulting Team first set out to verify whether the data received was clean and if it corresponded to the information provided in the briefing from the marketing department - both were quickly verified. The average number of impressions each consumer saw was 8 with a standard deviation of around 22. The proportion of customers in control and test who purchased was 49% and 51%, respectively. This was the first hint that the choice-based sample of 50/50 purchases vs non-purchasers may have been ill-founded. The internet browsing behavior across the control and test groups was indistinguishable which gave evidence that the experiment did not suffer from selection bias - t-tests also confirmed this assertion. Power tests revealed that Star Digital would be able to detect a 5.1% effect reliably with the sample provided and the minimum sample required to detect a 10% effect between control and test was 1237, roughly half the number of customers provided in the control sample. At this point, the experiment seemed to be free of endogeneity, selection bias, interference, and equipped with sufficient power.

To begin the causal analysis, the team studied whether there was a frequency effect of advertising on purchasing behavior since it would likely inform the other two key questions. The team bundled all impressions across the websites into one independent variable and leveraged a logistic regression to regress the probability of purchase on total impressions and test(0 or 1). Upon analyzing the results, it was found that treatment did improve the probability of purchase. However, the effect was not nearly as significant as total impressions' effect on the probability of purchase. The team then plotted test cases only containing the control group to see whether total impressions raised the probability of purchase substantially even when the ads were for charity. The team also assessed the change in probability of purchase when going from control to test on increasingly larger total impressions test cases. The results were disappointing and led the team to conclude that nothing assertive could be said about the effectiveness of this study due to omitted variable bias, poor experimental design, and limited statistical power. Star Digital should consider the following for future studies:

- Collect data regarding the amount of time individuals spend on the internet; it is impossible to gauge the true impact of the frequency of advertising without this information or more advanced observational econometrics
- Select a 50/50 control vs test sample instead of a customer sample of 50/50 purchased vs non-purchased
- Be more explicit about the desired minimum lift to inform the amount of data needed to reduce costs and to ensure meaningful conclusions can be made

0.0.2 Exploratory Analysis & Experimental Design Checks

```
library(tidyverse)
library(pwr)
library(aod)
library(naniar)

setwd("~/MSBA/spring/Econometrics/hw/1")
data <- read.csv('starDigital.csv')

data <- data %>% mutate(imp_total = sum1to5 + imp_6)

miss_var_summary(data)

table(data$purchase)
table(data$test)
```

Overall 50% of the sample purchased suscription packages and 89% of the sample belong to test group, conforming to the marketing department's intended experimental design. The data provided is also free of missing values.

```
test <- data %>%
  filter(test == 1)

# mean(test$imp_total)
# sd(test$imp_total)
round(table(test$purchase)[2]/(table(test$purchase)[1]+table(test$purchase)[2]),3)*100

##      1
## 50.5

control <- data %>%
  filter(test == 0)

# mean(control$imp_total)
# sd(control$imp_total)
round(table(control$purchase)[2]/(table(control$purchase)[1]+table(control$purchase)[2]),3)*100
```

```
## 1
## 48.6
```

51% of the test group and 49% of the control group purchased a Star Digital package, corresponding to the choice-based sample Star Digital intended to collect. This is the first hint at how this choice-based sample may be problematic since the purchasing behavior seems to be very similar between the control and test group. The distribution of impressions seen, irrespective of the ad type, indicating that the control and test groups have roughly the same internet browsing behavior.

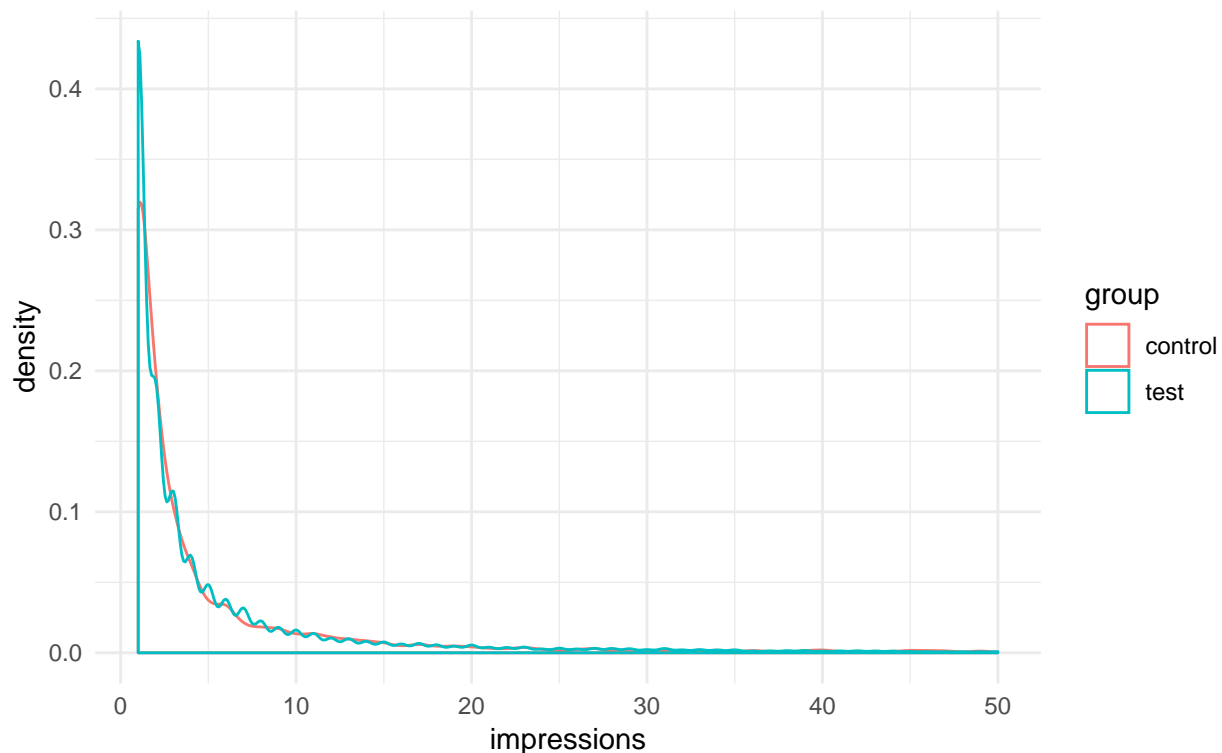
```
bb <- data %>%
  mutate(group = ifelse(test == 1, 'test', 'control')) %>%
  rename(impressions = imp_total)

bb$group <- as.factor(bb$group)

bb %>%
  filter(impressions < (mean(impressions) + 2 * sd(impressions))) %>%
  ggplot() + geom_density(aes(x = impressions, color = group)) + theme_minimal() +
  labs(title = 'Internet Browsing Behavior Between Control and Test is Randomized',
        subtitle = 'Figure 1')
```

Internet Browsing Behavior Between Control and Test is Randomized

Figure 1



Above is an illustration of the distributions of the amount of ads seen by customers in the control and test groups. This is of interest to Star Digital because the amount of ads seen by a specific customer is likely a function of the frequency at which that person browses the internet. If there was a substantial difference between the distributions above, there would be evidence against the validity of the control group since the internet browsing behavior must be randomized across test and control. This logic assumes that internet browsing behavior of an individual is time-invariant and could have been leveraged by Star Digital prior to the experiment.

To more precisely establish whether the internet browsing behavior was random across treatment and control, t-tests were conducted for the number of impressions in total, impressions on websites 1-5, and impressions on website 6.

```
t.test(imp_total ~ test, data = data)
t.test(sum1to5 ~ test, data = data)
t.test(imp_6 ~ test, data = data)
```

Conforming to figure 1, there was no statistical evidence that the internet browsing behavior of the test and control groups are demonstrably different across any of the websites with the lowest p-value of any of the tests being 0.66.

The team then sought to understand if there were any other factors in this study that may have caused endogeneity. There was no measurement of the TV viewership behavior or the prior familiarity of each customer to Star Digital's brand; both of these unobserved confounds could very well be correlated with the purchasing decision of a customer. However, the team has no evidence indicating that either confound is correlated with a customer's internet browsing behavior or likelihood to receive treatment. Thus, there is no evidence of omitted variable bias at this point. The team also did not obtain any reason to believe that this study was hindered by simultaneity, measurement error, or interference. Although, these should be rigorously considered by Star Digital's marketing team since they likely have more domain knowledge on the matter.

Now that threats to causal inference were understood for this experiment, the question became: was there enough data collected to detect the minimum lift of interest for Star Digital or was too much data collected? Below the team has conducted power tests to estimate the minimum number of customers that would be needed in control and test to detect a lift of 10% as well as the minimum lift this study could reliably detect given the sample size of the control and test groups - 2656 & 22647. It was assumed that a 90% degree of confidence that there was, in fact, significant results and an 80% degree of confidence that the study did not fail to detect a treatment effect when there was one would be reasonable for management in this case.

```
power.t.test(n=NULL, power=0.8, sig.level=0.1, delta=0.1)
```

```
##
##      Two-sample t test power calculation
##
##              n = 1237.188
##              delta = 0.1
##              sd = 1
##              sig.level = 0.1
##              power = 0.8
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

In order to reliably detect a 10% difference in the means across the control and test groups, Star Digital needed 1237 samples in both the control and test group given the parameter settings above.

```
pwr.t2n.test(n1 = 2656, n2 = 22647, sig.level = 0.1, power = .8)
```

```
##
##      t test power calculation
##
##              n1 = 2656
##              n2 = 22647
##              d = 0.05102679
##              sig.level = 0.1
##              power = 0.8
##      alternative = two.sided
```

With the amount of data at hand, the experiment will be able to identify a 5.1% difference in the means between the control group and test group. This is an important note for management since the analysis below will not be able to reliably detect a difference in the purchasing rate between control and test below 5.1%. This may be problematic if Star Digital's desired effect is below this threshold, but in general this is a pretty small effect, indicating that there is sufficient data for this study to provide reliable estimations. At this point, the team felt well prepared to begin answering the business questions of interest.

0.0.3 Is there a frequency effect of advertising on purchase?

The team has selected logistic regression as the modeling technique for this case since the dependent variable is binary. This does convolute the interpretation of the estimated coefficients for the independent variables but the team has taken care to provide an intuitive interpretation by deriving the `logit2prob` function below. This allowed the team to convert the coefficients and dependent variable estimates into a probability from the naturally provided log-odds estimates of the logit function. To ascertain the pure effect that frequency of advertising had on the probability of purchase for a customer the team bundled impressions across all sites into one independent variable as an input for the model below.

```
logit2prob <- function(logit){
  odds <- exp(logit)
  prob <- odds / (1 + odds)
  return(prob)
}
```

```
logit <- glm(purchase ~ imp_total + as.factor(test), data = data, family = "binomial")
```

```
summary(logit)
```

```
##
## Call:
## glm(formula = purchase ~ imp_total + as.factor(test), family = "binomial",
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.7408  -1.1238   0.1405   1.2133   1.2699
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.244063   0.040023  -6.098 1.07e-09 ***
## imp_total      0.029180   0.001294  22.551 < 2e-16 ***
## as.factor(test)1 0.072376   0.041697   1.736  0.0826 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 35077  on 25302  degrees of freedom
## Residual deviance: 34209  on 25300  degrees of freedom
## AIC: 34215
##
## Number of Fisher Scoring iterations: 5
```

The results above can be interpreted as follows:

- Each one-unit change in total impressions increases the log odds of purchase by 0.0292. The p-value indicates that it is very unlikely that repeated estimates would result in a test statistic at least as extreme as what was observed, given that the coefficient is in fact 0. Thus, there is strong evidence

to believe that there is a positive relationship between the number of impressions and the probability of purchase.

- Changing from control to test increases the log odds of purchase by 0.0724. The p-value indicates that there is an 8% chance a more extreme test statistic would be found in future estimates, given that the coefficient is in fact 0. Again, there is strong evidence to believe that showing customers Star Digital ads rather than charity ads increases the probability of purchase.

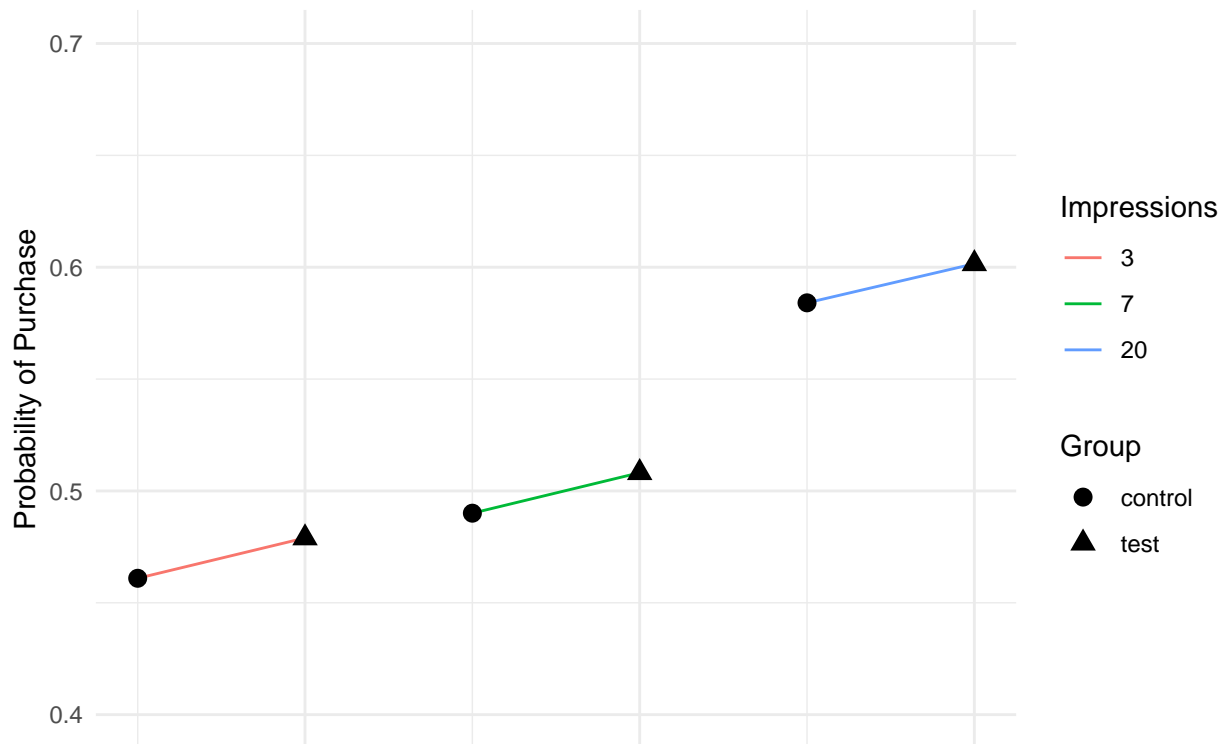
Now, the above interpretations are indeed convoluted, but the team intended to ensure that proper statistical rigor was applied throughout this study. Below the team has provided a few test cases that give a more intuitive understanding of the results. Based upon the above interpretations the higher the “case” number, the higher the probability of purchase should be for a given customer.

```
N = 6
imp_total <- c(3, 3, 7, 7, 20, 20)
test <- c(0, 1, 0, 1, 0, 1)
case <- c(1, 2, 3, 4, 5, 6)
df <- data.frame(imp_total, test, case)
```

```
proba <- c()
for (i in 1:N) {
  proba[i] <- logit2prob(predict(logit, df[i, 1:2]))
}
```

Test Group Has Higher Probability of Purchase

Figure 2



```
N = 6
imp_total <- c(3, 7, 17, 27, 37, 47)
test <- c(0, 0, 0, 0, 0, 0)
case <- c(1, 2, 3, 4, 5, 6)
df <- data.frame(imp_total, test, case)
```

```

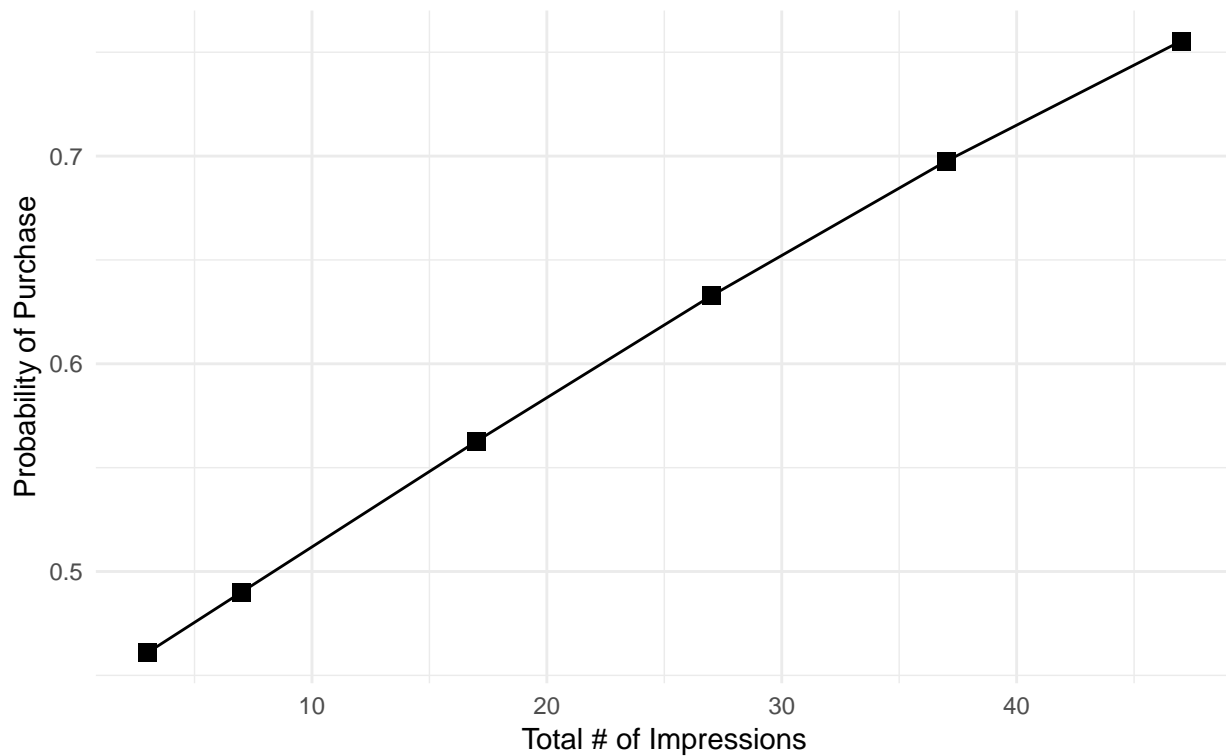
proba1 <- c()
for (i in 1:N) {
  proba1[i] <- logit2prob(predict(logit, df[i, 1:2]))
}

control_df <- data.frame(proba1, imp_total)

ggplot(control_df) + geom_point(aes(x = imp_total, y = proba1), shape = 'square', size = 3) +
  geom_line(aes(x = imp_total, y = proba1)) +
  labs(title = 'Probability of Purchase Increases with # of Impressions Even in Control',
        subtitle = 'Figure 3') +
  ylab('Probability of Purchase') + xlab('Total # of Impressions') + theme_minimal()

```

Probability of Purchase Increases with # of Impressions Even in Control
Figure 3



```
round(((proba[2] - proba[1]) / proba[2]), 4) * 100
```

```
## [1] 3.76
```

```
round(((proba[4] - proba[3]) / proba[3]), 4) * 100
```

```
## [1] 3.69
```

```
round(((proba[6] - proba[5]) / proba[5]), 4) * 100
```

```
## [1] 2.99
```

The substantial linear rise in the probability of a consumer purchasing as the number of impressions goes up, even in the control group, is concerning. This phenomenon may be attributed to the unobserved confound “customers increasingly turning to the internet to consume media” mentioned at the beginning of Star Digital’s provided briefing. This could be causing omitted variable bias since it is entirely plausible that the more time

a customer spends online, the more likely that customer would be to consume video services via the internet and see a large volume of ads on various websites.

Furthermore, the effect on the probability of purchase when going from control to treatment only results in a 3.76% change in the expected probability of purchase for consumers at the lowest test case in terms of the number of impressions. The percentage change in the probability of purchase begins to diminish as the number of impressions goes up when going from control to test. Even the test case with the lowest number of impressions seems to produce an unreliable effect given that this experiment can only detect an effect of 5.1% or above with confidence.

0.0.4 Which site should Star Digital advertise on?

Given the findings above, nothing can be said with confidence as to whether the advertising campaign is producing the results that the marketing team set out to achieve, irrespective of the particular website.

0.0.5 Is online advertising effective for Star Digital?

Cost/Revenue info:

- Sites 1-5: \$25 per 1000 impressions
- Site 6: \$20 per 1000 impressions
- Purchase: \$1,200 in lifetime contribution

```
rev = data %>% mutate(cost = ((25 * sum1to5 / 1000) + (20 * imp_6 / 1000)),
  revenue = (1200 * purchase) - ((25 * sum1to5 / 1000) + (20 * imp_6 / 1000)))

rev %>% group_by(test) %>% summarise(avg_cost = mean(cost),
  avg_rev = mean(revenue), sd_rev = sd(revenue))

t.test(revenue ~ test, data = rev)
```

When comparing the revenues of the control and test groups, there did seem to be a substantial difference with a p-value of 0.0615. However, this does not mask the unreliable results found above as to the effectiveness of this advertising campaign.

0.0.6 Final Recommendations

The results obtained throughout this study are disappointing. Star Digital will not be able to make data-driven decisions during their next marketing budget allocation due to poor experimental design. For future studies, Star Digital should make a few changes to enhance their understanding of customer behavior by being more explicit about the minimum lift desired. This will allow for a precise estimate of the sample size needed in the control and test groups. It was estimated above that Star Digital only needed 1237 customers in the control and 1237 customers in the test group if their minimum lift was 10%. This is nearly half the customers that were included in the control group. By collaborating with the Carlson Consulting Team on this matter before the study, Star Digital could have saved a substantial amount of money on non-value-added charity ads. Most importantly, it was found that this study suffered from omitted variable bias as well as a choice-based sample that limited the statistical power of this analysis. This produced an irrational estimate of the coefficient for total impressions and allowed the team to only be able to detect a 5.1% effect or greater between control and test groups.

The issues mentioned above have resulted in inconclusive recommendations and Star Digital will likely have to redo the study. Beyond selecting a 50/50 control vs test sample instead of a sample of 50/50 purchased vs non-purchased for future studies, Star Digital may also want to consider collecting information as to when customers purchased during the two months and how much time a consumer spends on the internet in general. This information would provide a better understanding of which customers require fewer impressions to purchase and could be beneficial for targeting efforts and would also allow for the study to control for the amount of time a consumer spends on the internet.