Estimating Ad Impact on Clicker Conversions for Causal Attribution: A Potential Outcomes Approach*

Joel Barajas† Ram Akella‡ Aaron Flores§ Marius Holtan§

Abstract

We analyze the causal effect of online ads on the conversion probability of the users who click on the ad (clickers). We show that designing a randomized experiment to find this effect is infeasible, and propose a method to find the local effect on the clicker conversions. This method is developed in the Potential Outcomes causal model, via Principal Stratification to model non-ignorable post-treatment (or endogenous) variables such as user clicks, and is validated with simulated data. Based on two large-scale randomized experiments, performed for 7.16 million users and 22.7 million users to evaluate ad exposures, a pessimistic analysis for this effect shows a minimum increase of the campaigns effect on the clicker conversion probability of 75% with respect to the non-clickers. This finding contradicts a recent belief that clicks are not indicative of campaign success, and provides guidance in the user targeting task. In addition, we find a larger number of converting users attributed to the overall campaign than those attributed based on the click-to-conversion (C2C) standard business model. This evidence challenges the well-accepted belief that C2C attribution model over-estimates the value of the campaign.

1 Introduction and Motivation

Recent developments in online campaign attribution and evaluation have demonstrated the effectiveness of display advertising on user conversion and search keywords probabilities [17, 18, 5]. These findings have motivated advertisers and ad networks to measure the effectiveness of campaigns in metrics other than user clicks. The belief that user clicks are not informative to measure the success of a campaign is increasingly gaining acceptance in the research community and industry. Dalessandro *et al.* concluded that user clicks do not correlate with user conversions, and that user targeting based on clicks is statistically indistinguishable from random guessing [9]. These findings are drawn based on the power of user clicks to predict conversions in observational data. However, a significant percentage of these conversions are likely to be unrelated with, and not caused by, the campaign, as it is standard in online advertising attribution analysis [18].

A more accurate approach is to measure the campaign effect on the conversion probability of the users who click on the ad (clickers) with a randomized experiment. Based on this effect, we can determine the importance of the click in the user targeting optimization. However, to design such experiment one would need to randomize the users into control/study groups after finding the clickers. This randomized design is not feasible because, to observe the user selection introduced by the click event, the online ad must be displayed to the users of the study and the control groups. Ideally, this user selection needs to be known before the campaign or the placebo ad is displayed in order to randomize the clicking users properly.

We propose to find the local average campaign effect on the clicker conversions based on the standard campaign evaluation randomized design. To the best of our knowledge, the proposed method is the first approach in the ad effectiveness measurement literature that estimates this effect based on randomized experiments. In the context of the Potential Outcomes causal model, we use Principal Stratification [10] to condition the campaign effect on the user click event. This framework allows us to model the treatment causal effect conditional on post-treatment variables, which are affected by the treatment and consequently are non-ignorable [24]. We compare the effect on the conversion probability of the clickers and the non-clickers to determine if the click event provides significant information to separate users with higher or lower campaign impact.

We discuss related advertising evaluation literature and causal modeling of post-treatment variables literature in section 2. We approach the problem in two phases: the randomized design, and the causal modeling given this design. For the randomized design we

*This work is partially funded by CONACYT UC-MEXUS grant 194880, CITRIS and AOL Faculty Award.

†Electrical Engineering, University of California, Santa Cruz, CA, USA. Email: jbarajas@soe.ucsc.edu. Main contact.

‡School of Information, University of California, Berkeley, CA, USA. Email: akella@ischool.berkeley.edu. University of California, Santa Cruz, CA, USA. akella@soe.ucsc.edu

§AOL Research, Palo Alto CA, USA. Email: {aaron.flores, marius.holtan}@teamaol.com

discuss the issues that prevent us from designing an experiment focused solely on the clicking users in section 3.1. We illustrate the randomized design we employed, which is focused on the measurement of the ad exposure effectiveness. In the casual modeling phase, we propose a method in the Potential Outcomes causal model to estimate the campaign effect on the clicker conversions based on the randomized experiment of section 3.2. In section 4 we analyze the effectiveness of the estimation method with simulated data, and discuss the results for two large-scale randomized experiments in detailed. Finally, we discuss the impact and benefits of the estimation approach in the ad effectiveness literature, as well as the impact of the results for user targeting and attribution in section 5.

2 Related Work and Background

Online Display Advertising poses unique challenges to measure the causal effect of online campaigns on the probability of online sales (or conversions). These challenges include: small propensity of an online user to convert (typically in the order of 1 in 10,000 or less [5]), small average causal effects and campaign lifts, and a severe user selection bias introduced by sophisticated campaign management that targets users and executes the bidding policy in ad exchanges. Due to these constraints, the use of randomized experiments is becoming the standard evaluation practice [8, 18, 28, 6], as opposed to the analysis of observational studies [7, 15, 19, 27, 4] where no randomization is deployed. Overall, observational studies are likely to over-estimate the campaign effectiveness [18]. Lewis and Reiley model the campaign effect on clicker conversions by analyzing observational data in a differences-in-differences approach, in spite of the availability of randomized data [17]. Dalessandro *et al.* performed an analysis to asses the effectiveness of optimizing user clicks in user targeting based on observational data, which tends to find correlations between user clicks and a significant amount of conversions not caused by the campaign [9]. Even when randomized experiments are performed, the user clicks are often discarded [16] due to the lack of effective techniques to incorporate them in the causal analysis.

In the Potential Outcomes causal model, Principal Stratification has been successfully used to find treatment effect when a selection bias is unavoidable in a randomized experiment [10]. The analysis of randomized experiments with non-compliance, where the randomly assigned individuals might opt out of the experiment due to treatment side effect [13, 14], is one of the most successful applications. Similarly, the analysis of right-censored data due to non-ignorable individual death [25, 11], and the education programs assessment with truncated data due to student drop-out [29], are other problems where an unavoidable (or endogenous) bias is addressed by Principal Stratification. A different approach when *intermediate* variables, such as user clicks, are observed in the "causal path" is to consider "causal mediation" or indirect effects [22, 21]. However, Rubin illustrates the risk of a bias analysis when the potential outcomes are not properly modeled by indirect effect analysis [23].

We propose a method to model the causal effect of advertising in the sub-population of clickers and non-clickers. This approach closes the gap between a purely observational analysis of the effect on the clicker conversions and the analysis of this effect with randomized experiments. As the problem is different than previously addressed problems by Principal Stratification, we solve the identifiability problem, typical in these problems, with mild and reasonable assumptions in online advertising. The uncertainty of the estimations is modeled in a Bayesian framework and a Gibbs-sampling based inference approach.

3 Proposed Methods

3.1 Randomized Design The current practice to estimate the campaign causal effect is to run a randomized experiment assuming the ad creative is the *treatment* to evaluate. In this context, the online visiting users are randomly assigned to the control or the study groups before the campaign starts. These users are maintained in the assigned group during the entire duration of the campaign. For those assigned to the study group, the campaign ad is displayed, while a placebo ad (assumed to be completely unrelated to the advertiser running the campaign) is displayed to the users of the control group [18, 28]. Then, the online users are tracked, based on tracking cookies or e-mail sign-ups, to observe if they convert in the advertiser website or not. In practice, the placebo ads are displayed by running a placebo campaign, which replicates the user selection (or targeting) performed by the advertising campaign.

Following a similar logic, to design a randomized experiment to estimate the campaign effect on the clicker conversions one can run a placebo campaign to replicate the clicking user selection. Then, a placebo ad would be displayed to the users in the control group once this selection is observed to the placebo campaign. Unfortunately, running such a design is not feasible because the clicking user population segment cannot be observed without showing the campaign ad. This prevents us from running a randomized experiment focused on the sub-population of clicking users.

To avoid relying on fully observational data, whose effectiveness to find causal estimates has been seriously

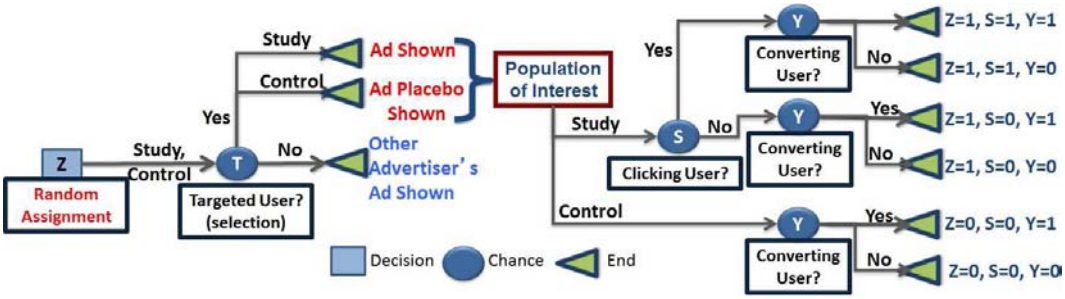


Figure 1: Randomized Design. The user clicks and conversions are collected for the user population of interest.

questioned in Online Advertising [18], we take advantage of the randomized design used in standard campaign evaluation. Thus, we randomly assign the users to control and study groups and focus on those selected by the ad-network targeting engine. This user population represents the universe of users for the effects of this paper. Fig 1 illustrates the randomized design. As a consequence of this design, the user selection introduced by the user clicks becomes a post-treatment variable, or a random variable that is affected by the campaign [10]. In the Potential Outcomes causal model, this variable is non-ignorable and must be modeled to find causal campaign effect on these user sub-populations [24][1].

3.2 Campaign Causal Estimation

3.2.1 Potential Outcomes and Principal Stratification Potential Outcomes Causal Model, also known as Rubin Causal Model (RCM)[24], is based on the analysis of units, treatments and potential outcomes. Fundamentally, RCM analyzes the unit potential outcomes to each of the treatments. For two treatment arms, control and study, this framework implies that half of the data is missing. This is because we can never observe the response of a unit in both arms. Thus, the causal inference problem is addressed as a missing-value inference problem. This problem is generally approached with a Bayesian parametric model to find the mean posterior predictive distribution. RCM incorporates the treatment assignment mechanism to offer a clear distinction between randomized experiments and observational studies. The key requirement for randomized experiments to be unbiased is that the treatment assignment mechanism must be *ignorable*[24]. This condition implies that this assignment does not provide any information in the causal inference problem. Standard notation in RCM is to consider the variable of a user i for a given treatment arm Z_i as $Y_i(Z_i)$. In spite of the ability to model the treatment effect on post-treatment variables S, typically the main interest is to estimate the treatment effect on Y conditional on S. However, this is not straightforward because $S_i(0) \neq S_i(1)$, and consequently S_i is not ignorable. Therefore, conditioning the effect estimates on the observed values of S introduces a post-treatment bias.

Principal Stratification modeling provides a framework to estimate unbiased treatment effect conditional on post-treatment variables [10]. The key component in this framework is to define user classes, or strata, S_i^P which are invariant to the treatment assignment. Thus, the probability of S_i^P must be independent (or ignorable) of the treatment assignment Z_i, to enforce that no treatment effect on the strata is allowed in the inference process.

3.2.2 Campaign Causal Effect on the Clicker Conversions The randomized design of Fig 1 allows us to record the user clicks for both treatment groups. However, a user click on the campaign ad is not comparable to a click on a placebo ad. As a result, the user selection made by the clicker indicator in the study group is missing in the control group.

We define the following indicator random variables for each user i: Z_i for control/study group user assignments $\{0,1\}$, S_i for non-clicker/clicker users $\{0,1\}$, Y_i for non-converting/converting users $\{0,1\}$. Given that the user must be in the study group to click the ad, the users of the control group never click the ad and consequently $\mathrm{S}_i = 0$ for these users. Therefore, we define the principal strata S_i^P, or user classes C_i as follows:

$$S_i^P = \left\{ \begin{pmatrix} S_i(0) \\ S_i(1) \end{pmatrix} \right\} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, \tag{3.1}$$

$$C_i = \begin{cases} 0 & \text{if } S_i^P = (0,0)' \\ 1 & \text{if } S_i^P = (0,1)' \end{cases}$$

Table 1 illustrates the observed and missed data in the RCM notation. $\mathrm{C}_i = 1$ are the users who click on the ad when they are assigned to the study group (clicker-

[1] In econometric causality, the user clicking indicator would be endogenous because this variable is not controllable by the experimenter [12].

Table 1: User counts based on the user potential outcomes. N^y_{cz}, where $C_i = c$, $Z_i = z$, $Y_i = y$, are user counts for the given values of Y, Z, C. Missing values are presented as *.

User Counts	**Potential Outcomes**				**Treatment Assignment**	**Principal Stratum**	
	Control		Study				
N^y_{cz}	$S_i(0)$	$Y_i(0)$	$S_i(1)$	$Y_i(1)$	Z_i	$(S_i(0), S_i(1))$	C_i
$N^0_{\{0,1\}0}$	0	0	*	*	0	(0,*)	*
$N^1_{\{0,1\}0}$	0	1	*	*	0	(0,*)	*
N^0_{01}	0	*	0	0	1	(0,0)	0
N^1_{01}	0	*	0	1	1	(0,0)	0
N^0_{11}	0	*	1	0	1	(0,1)	1
N^1_{11}	0	*	1	1	1	(0,1)	1

if-assigned).$C_i = 0$ are the users who do not click on the ad regardless of the treatment group they are assigned to (never-clickers). Based on these definitions, we let C_i to be Bernoulli distributed with parameter π, and Y_i to be Bernoulli distributed with parameters θ_{cz} for the 4 combinations $C_i = c$, $Z_i = z$. Assuming a Bayesian approach in the parameter estimation, we define $\Theta = \{\theta_{cz}, \pi\}$ as random variables.

Similar to the case of randomized experiments with non-compliance [13, 2], this model is not identifiable if no further constraints are imposed. To estimate the campaign effect on the clicker conversions, we observe the stratum indicator C_i and the conversion indicator Y_i for the users in the study group $Z_i = 1$. These observed indicators allow us to estimate the user conversion probability for both strata in the study arm without constraints. By randomization, we know that the probability of observing this user selection π is the same in both treatment groups, which follows from the definition of the principal strata [10]. However, the user conversion probability for both principal strata users in the control group $\{\theta_{10}, \theta_{00}\}$ are not identifiable. Thus, to guarantee the model is identifiable we assume positive campaign effect. This assumption translates into $\theta_{c1} \geq \theta_{c0}$ for $c = \{0, 1\}$. Therefore, letting the indicator function be $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise, and assuming a prior distribution $P(\Theta)$, we have the joint distribution:

$$P(Y, Z, D, \Theta) = P(\Theta) I_{[0,\theta_{01})}(\theta_{00}) I_{[0,\theta_{11})}(\theta_{10}) \times \prod_{\forall i} P(C_i|\pi) P(Z_i) P(Y_i|C_i, Z_i, \theta_{cz}) \tag{3.2}$$

We assume standard conjugate Beta prior distributions for the Bernoulli distributed random variables $\Theta = \{\theta_{cz}, \pi\}$ for $c = \{0, 1\}$ and $z = \{0, 1\}$. For numerical stability, we use the Jeffreys prior distribution, Beta(0.5.0.5), which assumes a prior sample size of 1. We experiment with different prior probability but with the same sample size of 1. Given the number of users employed to estimate the conversion probabilities θ_{cz}, the effect of these prior probabilities becomes negligible.

We note that Balke and Pearl have reported in the context of imperfect compliance that estimating these effects is not feasible with no constraints [2]. They provide a set of bounds based on a method of moments assuming a large sample of individuals. The model of Eq 3.2 is fitted without relying on those bounds as a consequence of: the positive effect constraint, and the full observance of the potential outcomes for the users of the study group detailed above.

3.2.3 Model Estimation The inference objective of the joint distribution of Eq 3.2 is to find the posterior distribution of the parameters Θ given the observed data from Table 1. We denote the set of observed counts as D_{obs}. We solve this inference problem using Gibbs sampling by sampling from the conditional posterior distributions. Given an initial guess for Θ^0 and similar to standard mixture methods, we sample the missing user clicking indicator for the users in the control group and estimate the counts N^y_{c0} for $c = \{0, 1\}, y = \{0, 1\}$. We perform this sampling step based on the probability of user clicking assignment, C^y_{i0}. We denote these sampled counts as $D_{samp} = \{N^y_{c0}\}$ for $c = \{0, 1\}, y = \{0, 1\}$. Given the augmented user counts, $\{D_{obs}, D_{samp}\}$, we sample the parameters Θ. The sampling distributions of the user conversion probabilities for the study group and the probability of a clicking user, $\{\theta_{c1}, \pi\}, c = \{0, 1\}$ are Beta distributions. For the constrained parameters, $\{\theta_{c0}\}, c = \{0, 1\}$, the conditional posterior distributions become Beta distributions truncated to be nonzero at the range $[0, \theta_{c1})$ for $c = \{0, 1\}$. We sample

Algorithm 1 Gibbs Sampling Algorithm based on the joint distribution of Eq. 3.2

Define $D_{obs} = \left\{ N^y_{c1}, N^y_{\{0,1\}0} \right\}$ for $c = \{0,1\}, y = \{0,1\}$ from Table 1
Define $D_{samp} = \{N^y_{c0}\}$ for $c = \{0,1\}, y = \{0,1\}$
Set $\alpha_0 = 0.5$
Initial guess $\Theta^0 = \{\theta_{cz}, \pi\}^0$, for $c = \{0,1\}, z = \{0,1\}$
for $s \leftarrow 1$ to $N_{burnin} + N_{samples}$ **do**
  Set $P(C^y_{i0} = 1|\Theta, D_{obs}) = \dfrac{\pi(\theta_{10})^y(1-\theta_{10})^{(1-y)}}{\pi(\theta_{10})^y(1-\theta_{10})^{(1-y)} + (1-\pi)(\theta_{00})^y(1-\theta_{00})^{(1-y)}}, \quad y = \{0,1\}$
  Draw $N^y_{10}|\Theta, D_{obs} \sim \text{Binomial}\Big(N^y_{\{0,1\}0}, P(C^y_{i0} = 1|\Theta, D_{obs})\Big), \quad y = \{0,1\}$
  Set $N^y_{00} = N^y_{\{0,1\}0} - N^y_{10}, \quad y = \{0,1\}$
  Draw $\theta^s_{c1}|\Theta_{-\theta_{c1}}, D_{samp}, D_{obs} \sim \text{Beta}\left(\alpha_0 + N^1_{c1}, \alpha_0 + N^0_{c1}\right), \quad c = \{0,1\}$
  Draw $\theta^s_{c0}|\Theta_{-\theta_{c0}}, D_{samp}, D_{obs} \sim \text{Beta}\left(\alpha_0 + N^1_{c1}, \alpha_0 + N^0_{c1}\right) I_{[0,\theta_{c1})}(\theta_{c0}), \quad c = \{0,1\}$
  Draw $\pi^s|\Theta_{-\pi}, D_{samp}, D_{obs} \sim \text{Beta}\Big(\alpha_0 + \sum_{z,y} N^y_{1z}, \alpha_0 + \sum_{z,y} N^y_{0z}\Big)$
end for
Discard $\Theta^{1:N_{burnin}}$ and keep $\Theta^{N_{burnin}+1:N_{samples}}$

from a truncated Beta distribution using the method detailed at [20]. This sampling process is repeated for $N_{burnin} + N_{samples}$. After discarding a set of burn-in samples, N_{burnin}, a set of random samples of the posterior distribution is obtained, $\Theta^{1:Nsamples}$. Algorithm 1 illustrates this sampling process and the posterior distribution expressions.

This inference procedure allows us to estimate the variability (or heterogeneity) of the local campaign effect from the posterior random set of samples $\Theta^{1:N_{samples}}$ Thus, the local average campaign effect on the clicker, LATE_{Click}, and non-clickers, $\text{LATE}_{NoClick}$, conversions are estimated based on these posterior distribution samples as follows:

$$\begin{aligned} LATE_{Click} &= E[Y_i|C_i = 1, Z_i = 1, \theta_{11}] \\ &\quad - E[Y_i|C_i = 1, Z_i = 0, \theta_{10}] = \theta_{11} - \theta_{10} \\ LATE_{NoClick} &= E[Y_i|C_i = 0, Z_i = 1, \theta_{01}] \\ &\quad - E[Y_i|C_i = 0, Z_i = 0, \theta_{00}] = \theta_{01} - \theta_{00} \end{aligned} \tag{3.3}$$

Therefore, LATE_{Click} and $\text{LATE}_{NonClick}$ become random variables allowing us to estimate their posterior confidence interval. In addition, we find the lifts by finding the ad exposure effect with respect to the conversion rate in the control group for both populations.

We estimate the proportion of attributed converting users for these sub-populations with respect to the converting users in the study group (ATRB_{Click}, $\text{ATRB}_{NoClick}$) as follows:

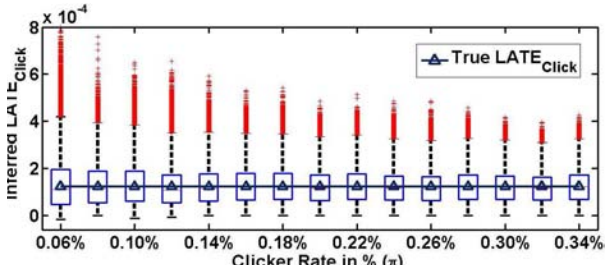
$$\begin{aligned} ATRB_{Click} &= LATE_{Click} \\ &\quad \times \left(N^0_{11} + N^1_{11}\right) / \left(N^1_{01} + N^1_{11}\right) \\ ATRB_{NoClick} &= LATE_{NoClick} \\ &\quad \times \left(N^0_{01} + N^1_{01}\right) / \left(N^1_{01} + N^1_{11}\right) \end{aligned} \tag{3.4}$$

These metrics provide the campaign value in terms of attributed converting users, based on the campaign effect per user and the size of the population.

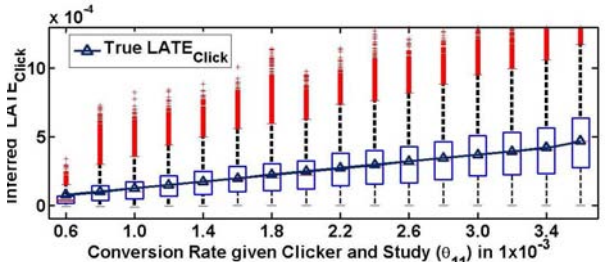
4 Results

4.1 Validation One of the main challenges to analyze the campaign effect on the clicker conversions is the low probability of clickers. Lewis *et al.* reported a clicker rate of 0.254% in a large-scale online experiment for more than 35 million users [18]. Even sparser is the probability of clicker and converter. In a exploratory campaign, where the user targeting is not optimized, we collect only eight clickers and converters out of more than 11 million users in the study group, which gives a 7.1e-7 joint probability of clickers and converters. Therefore, data sparsity prevents us from using large-sample approximations such as those in [2, 29] or those based on Normal approximations [12, 18], and consequently large confidence intervals are expected.

To analyze the power of the method to detect a given local campaign lift in the clickers, we assume a set of parameters $\{\Theta, P(Z_i), N_T\}$ (N_T is the total number of users) and simulate the data counts of Table 1. For each parameter set, we randomly generate 100 sampled data count sets. Then for each set, we run the inference Algorithm 1, where $N_{burnin} = 200$, $N_{samples} = 3,000$. Finally, we concatenate the posterior samples and obtain $\Theta^{1:100\times N_{samples}}$ and find $\{\text{LATE}_{Click}\}^{1:100\times N_{samples}}$ from Eq 3.4. Fig 2 shows the simulation results and the parameter values assumed as a function of: (a) the probability of a clicking user π, (b) the probability of a converting user in the study group θ_{11}. We expect that a successful campaign in optimizing clicks would increase these two parameters.



(a)



(b)

Figure 2: Boxplot of the posterior distribution of LATE_{Click} based on simulated experiments as a function of: (a) the clicker rate, (b) conversion rate of the clickers in the study group[3]. Assumed parameter values for: (a) {N_T =23e6, $P(Z)$ =0.5, θ_{11} =1e-3, θ_{01} =9.82e-5, lift LATE_{Click} =0.14, lift $\text{LATE}_{NoClick}$ =0.14}, (b) {N_T =23e6, $P(Z)$ =0.5, π =6.8e-4, θ_{01} =9.82e-5, lift LATE_{Click} =0.14, lift $\text{LATE}_{NoClick}$ =0.14}

In general, results from these simulated data experiments show that even when we do not observe the clicking user selection of the study group in the control group, we are able to infer the campaign effect in this sub-population without any bias. The only necessary assumption is to consider campaign positive effects, in spite of the low clicker rate. Both Fig 2(a)-(b) show a skew distribution. However, as we increase the clicker rate π in Fig 2(a), the posterior distribution of LATE_{Click} concentrates more at the true LATE_{Click}. This analysis shows that for a reasonable clicker rate of $\pi = 0.20\%$ or higher the effect distribution shows an increasingly well-defined peak. Fig 2 (b) shows that as the conversion rate of the clickers increases, and consequently the effect LATE_{Click} increases too assuming a fixed lift, the skewness level decreases, even though the clicker rate is low $\pi = 0.068\%$. However, the confidence interval is large due to the low clicker rate assumed. Overall, the clicker rate parameter shows to have higher impact on the estimator power than the conversion rate of the clickers.

[3]The zero effect appears to be in the intervals because the boxplot function obtains them based on a Normal approximation. Clearly the zero effect is not in the distribution as this is a constraint of Eq 3.2.

Table 2: Campaign data based on notation of Table 1. Campaign active duration for, Campaign 1: 16 days, Campaign 2: 28 days. $\{0,1\}$ represents unobserved clicking user indicator.

Count	Campaign 1	Campaign 2
$N^0_{\{0,1\}0}$	3,621,409	11,431,495
$N^1_{\{0,1\}0}$	314	961
N^0_{01}	3,535,571	11,328,649
N^1_{01}	347	1,014
N^0_{11}	2,414	9,799
N^1_{11}	2	8

We note that the lifts for LATE_{Click} and $\text{LATE}_{NoClick}$ are relative measures to the base conversion rates, which are different between clickers and non-clickers. In terms of campaign attribution, the relevant measurements are LATE_{Click} and $\text{LATE}_{NoClick}$. Therefore, although the lifts are equivalent for the experiments of Fig 2, different values of LATE_{Click} and $\text{LATE}_{NoClick}$ are tested assuming different campaign attribution for these user populations.

4.2 Randomized Experiment Data Description

We ran two large-scale randomized experiments at the *Advertising.com* ad network collaboratively with one advertiser in the financial information services sector. We randomly assigned the users to control/study groups based on the timestamp the tracking cookie was born. To avoid user contamination, we focused on the users whose cookie was born before the campaign started. Then, for each user visit to the set of publishers' websites, the targeting engine selected those users eligible to see the ad. After this selection was made, the campaign ad was displayed to the users in the study group, and a charity ad (placebo ad) was displayed to the users in the control group. Then, the users were tracked, based on their unique cookie, to observe the user clicks on the ad, and the user conversion at the advertiser's website. Fig 1 illustrates the randomized design, and Table 2 shows the aggregated user counts collected for these experiments based on the notation of Table 1. Although the advertiser might run the placebo and ad campaigns independently with the same targeting setup, the user randomization needs to be performed by the ad network to guarantee that no campaign or placebo ad is displayed to the incorrect group.

Both campaigns were run in a cost-per-thousand

(CPM) business model. Thus, the user targeting was not as optimized as in the case of conversion based attribution campaigns. The objective of the experiment is to perform an exploratory analysis of the campaign effectiveness with a limited budget, before the campaign is fully deployed. This practice is standard in campaign budget allocation [8]. The campaigns were run during different time periods. Although the same advertiser ran them, they were not related in any other aspect. For privacy reasons, we are not allowed to disclose the ad content or the advertiser identity.

4.3 Campaign Evaluation Results For comparison purposes, we estimate the ad click effect assuming we do not observe the control group of users, ATE^{obs}_{Click}, and the ad click effect with post-treatment bias, $\text{ATE}^{post}_{Click}$, which are defined by Eq 4.5.

$$\begin{aligned} ATE^{obs}_{Click} &= E[Y_i|S_i(1)=1, Z_i=1] \\ &\quad -E[Y_i|S_i(1)=0, Z_i=1] \\ ATE^{post}_{Click} &= E[Y_i|S_i(1)=1, Z_i=1] \\ &\quad -E[Y_i|S_i(0)=0, Z_i=0] \end{aligned} \tag{4.5}$$

ATE^{obs}_{Click} provides the conversion probability change of the clickers versus the non-clickers, which is an intuitive measurement of the value of the click indicator. In a pure prediction optimization framework, this measurement would represent the importance of the user click indicator feature. $\text{ATE}^{post}_{Click}$ represents the campaign effect conditional on the user clicking indicator without correcting for post-treatment bias introduced by this indicator, or its endogeneity.

Table 3 shows the effect results for the overall campaign, ATE_{Camp} obtained based on the user counts aggregated for the entire campaign, and the disaggregate effects $\text{LATE}_{NoClick}$, LATE_{Click} as defined by Eq 3.4. We set $N_{burnin} = 2,000$, $N_{samples} = 20,000$ for the Gibbs sampling of Algorithm 1. In addition, the non-causal estimates, ATE^{obs}_{Click} and $\text{ATE}^{post}_{Click}$, are depicted.

By inspecting the data of Table 2, the conversion rate for the clickers in the study group is close to 10 times higher, $N^1_{11}/(N^1_{11} + N^0_{11})$=8.27e-4 than for the non-clickers $N^1_{01}/(N^1_{01} + N^0_{01})$=9.81e-5 for campaign 1, and 8.16e-4 with respect to 8.95e-5 for campaign 2. ATE^{obs}_{Click} shows this (naive) effect estimation, based on a two-sample t-test with different variances and the Normal approximation. By looking at lift ATE^{obs}_{Click}, the upper bounds are larger than 1,000%, and the lower bound for campaign 2 is close to 300%. This result shows the over-estimation of the value of user clicks when the objective is to optimize conversion prediction. Even when the control group is used but the post-treatment bias is not corrected, represented by $\text{ATE}^{post}_{Click}$, the campaign effect is over-estimated severely. This over-estimation is very close in magnitude to the naive estimation when the control group is not used in the estimation (in average: lift ATE^{obs}_{Click}=743% vs lift $\text{ATE}^{post}_{Click}$=855% for campaign 1, and lift ATE^{obs}_{Click}=811% vs lift $\text{ATE}^{post}_{Click}$=870% for campaign 2). Therefore, not correcting the post-treatment bias eliminates most of the power of the randomized experiment to find the campaign effect on the clicker conversions.

We observe a clicker rate of less than 0.1%, which is a consequence of non-optimized campaigns. As a result, the campaign average effect, ATE_{Camp}, and the local average effect in the non-clickers, $\text{LATE}_{NoClick}$, are similar. This is because the vast majority of the users are non-clickers $C_i = 0$. We observe a larger effect for the clickers than for the non-clickers. As we discussed in section 4.1, we expect a skewed posterior distribution given the observed clicker rate. This skewness is more evident in the lift percentage estimation where the right-hand tail is in the order of hundreds. In spite of the wide confidence interval, we observe larger campaign effect in the users who click on the ad by analyzing the lower quantile of LATE_{Click} and the upper quantile of $\text{LATE}_{NoClick}$ for both campaigns. Therefore, a pessimistic scenario for the campaign effect on the clicker conversions shows an increase of **75%** (3.50e-5 - 2.00e-5 with respect to 2.00e-5) for campaign 1, and **252%** (4.30e-5 - 1.22e-5 with respect to 1.22e-5) for campaign 2, with respect to the campaign effect on the non-clicker conversions. This analysis shows that, as intuition suggests, user click probability is a measure of campaign success, and user clicks on ads are **not** random events as previous literature suggests [9].

In terms of the overall campaign attribution, we note that a significant amount of conversions attributed to the campaign is obtained from the non-clickers due to the user volume of this sub-population. For campaign 1, only 2.63% of the campaign attribution is attributed to the clickers (ATRB_{Click}=0.32 with respect to ATRB_{Camp}=12.17); and for campaign 2, 6.62% of the campaign attribution is attributed to the clickers (ATRB_{Click}=0.45 with respect to ATRB_{Camp}=6.80). Even when the effect on the clicker conversion probability is 252% higher than for the non-clickers, the volume of non-clickers is more than 900 times higher ($\pi/(1-\pi)$ =0.999/1e-3). Therefore, although the user click indicator is a measure of success, a pure click-to-conversion (C2C) attribution framework, where a user conversion is attributed to the last user click, tends to under-estimate the true campaign value. We estimate the overall median campaign attribution ATRB_{Camp} to be **12.17%** for campaign 1 and **6.8%** for cam-

Table 3: Average campaign effect and attributed conversion percentage respect to the number of converting users in the study group. Results obtained based on the data of Table 2. {*Low, Med, High*} are the {0.05, 0.5, 0.95} quantiles. C2C ATRB is estimated as the ratio of clickers and converters respect to the converting users in the study group: $N_{11}^1/(N_{01}^1 + N_{11}^1)$.

	Campaign 1				**Campaign 2**			
Measure		Low	Med	High		Low	Med	High
Clicker Rate, π	(%)	0.0667	0.0683	0.0699	(%)	0.0851	0.0865	0.0880
ATE_{Click}^{obs} (naive)	(1e-4)	-2.33	7.30	16.92	(1e-4)	2.54	7.26	12.00
lift ATE_{Click}^{obs}	(%)	-237	**743**	1720	(%)	282	**811**	1340
$\text{ATE}_{Click}^{post}$ (biased)	(1e-4)	-2.21	7.41	17.04	(1e-4)	2.54	7.26	12.00
lift $\text{ATE}_{Click}^{post}$	(%)	-255	**855**	1960	(%)	306	**870**	**1400**
ATE_{Camp}	(1e-5)	0.37	1.20	1.99	(1e-6)	-0.33	6.13	12.50
lift ATE_{Camp}	(%)	4.06	13.94	24.02	(%)	-0.38	7.30	15.43
ATRB_{Camp}	(%)	3.74	**12.17**	20.20	(%)	-0.37	**6.80**	13.87
$\text{LATE}_{NoClick}$	(1e-5)	0.34	1.16	**2.00**	(1e-6)	0.89	5.82	**12.20**
lift $\text{LATE}_{NoClick}$	(%)	3.81	13.46	24.21	(%)	1.04	6.97	25.13
$\text{ATRB}_{NoClick}$	(%)	3.48	11.78	20.22	(%)	0.99	6.45	13.53
LATE_{Click}	(1e-4)	**0.35**	4.61	13.72	(1e-4)	**0.43**	4.65	11.04
lift LATE_{Click}	(%)	7.28	150.77	874.20	(%)	7.21	145.85	813.12
ATRB_{Click}	(%)	0.02	**0.32**	0.95	(%)	0.04	**0.45**	1.06
C2C ATRB	(%)	-	**0.57**	-	(%)	-	**0.78**	-

paign 2 respect to the converting users in the study group ($N_{01}^1 + N_{11}^1$). However, the C2C attribution (C2C ATRB) percentage, % $N_{11}^1/(N_{01}^1 + N_{11}^1)$, is **0.57%** for campaign 1, and **0.78%** for campaign 2.

5 Impact and Conclusion

We have proposed a method to find the local advertising effect on the conversion probability of the users who click on the ad based on randomized experiments. We have shown that by not correcting the post-treatment bias based on the proposed method, the ad effect on the clicker conversions is as biased as the naive conversion rate lift between clickers and non-clickers. The main limitation of this method is that the average campaign effect must not be negative.

Contrary to the general belief that clicks are not indicative of campaign success, we have found evidence suggesting a higher campaign effect on the user conversion probability for those who click on the ad. In spite of the large confidence interval, a pessimistic analysis shows a substantial increase in the conversion probability for the clickers when compared to the non-clickers. These results are consistent with the largely used business models based on clicks. However, a significant percentage of the campaign attribution is observed in the non-clicking population, which is a drawback of the click-based attribution models. Surprisingly, by comparing the attribution results to the overall campaign with the C2C attribution, we have found that C2C under-estimates the causal attribution. This finding contradicts the general belief in the advertising industry that C2C conversion attribution model tends to overestimate the value of the campaign [1, 26]. We conclude that the population of clicking users is definitely more valuable than the non-clicking population. In terms of targeted advertising, the empirical results show a correlation between user clicks and ad causal effect on user conversions. Consequently, optimizing user clicks, which are less sparse than user conversions, optimizes the causally generated conversions by the ad exposure. However, the targeting policy should not optimize user clicks only, as a large percentage of users affected by the ad do not click on it. A combined policy to target clickers and non-clickers should be considered.

The method we have proposed opens a path for more studies of the user clicks with randomized experiments. The current study is limited to two campaigns with no attempt to optimize the user clicks. Finding the ad causal effect on the clicker conversions for C2C-optimized campaigns should increase the click probability and decrease the confidence intervals of the attribution. Similarly, recent evidence suggests different ad exposure effects between conversion-optimized and CPM campaigns [3]. Moreover, a very different conclusion would be drawn when ads that generate clicks by confusing the user are tested. Often the randomized data and the user clicks are available. However, this information is discarded under the assumption that no

relevant user intent information is being revealed, and the fact that reliable techniques to analyze this data are not available. Understanding what motivates a click, and why many users who are affected by the campaign do not click on the ad, is an open research problem that can be now analyzed with data from a randomized experiment.

References

[1] Atlas. Engagement mapping: A new measurement standard is emerging for advertisers, 2008.

[2] A. Balke and J. Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92:1171–1176, 1997.

[3] J. Barajas, R. Akella, M. Holtan, and A. Flores. Experimental designs and estimation for online display advertising attribution in marketplaces. *Submitted*, .

[4] J. Barajas, R. Akella, M. Holtan, J. Kwon, A. Flores, and V. Andrei. Dynamic effects of ad impressions on commercial actions in display advertising. In *Proceedings of 21st CIKM*, pages 1747–1751. ACM, 2012.

[5] J. Barajas, R. Akella, M. Holtan, J. Kwon, A. Flores, and V. Andrei. Marketing campaign evaluation in targeted display advertising. In *ADKDD '12*, pages 1–7. ACM, 2012.

[6] J. Barajas, R. Akella, M. Holtan, J. Kwon, A. Flores, and V. Andrei. Dynamic evaluation of online display advertising with randomized experiments: An aggregated approach. In *Proceedings ot WWW Conference*, Companion, pages 115–116, 2013.

[7] D. Chan, R. Ge, O. Gershony, T. Hesterberg, and D. Lambert. Evaluating online ad campaigns in a pipeline: causal models at scale. In *Proceedings of the 16th ACM SIGKDD*, pages 7–16. ACM, 2010.

[8] A. Chittilappilly. Using experiment design to build confidence in your attribution model. Online Metrics Insider, July 2012.

[9] B. Dalessandro, R. Hook, C. Perlich, and F. Provost. Evaluating and optimizing online advertising: Forget the click, but there are good proxies. *NYU Working Paper No CBA-12-02*, 2012.

[10] C. E. Frangakis and D. B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, March 2002.

[11] C. E. Frangakis, D. B. Rubin, M.-W. An, and E. MacKenzie. Principal stratification designs to estimate input data missing due to death. *Biometrics*, 63(3):641–649, 2007.

[12] J. J. Heckman. Econometric causality. Working Paper 13934, National Bureau of Economic Research, April 2008.

[13] G. W. Imbens and D. B. Rubin. Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25(1):305–327, 1997.

[14] H. Jin and D. B. Rubin. Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association*, 103(481):101–111, 2008.

[15] P. Kireyev, K. Pauwels, and S. Gupta. Do display ads influence search?: Attribution and dynamics in online advertising, 2013. Working paper (Harvard Business School).

[16] A. Lambrecht and C. Tucker. When does retargeting work? information specificity in online advertising. *Journal of Marketing Research*, 50(5):561–576, 2013.

[17] R. Lewis and D. Reiley. Online ads and offline sales: Measuring the effects of retail advertising via a controlled experiment on Yahoo!, August 2013. Working Paper.

[18] R. A. Lewis, J. M. Rao, and D. H. Reiley. Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of WWW Conference*, pages 157–166, 2011.

[19] P. Manchanda, J. P. Dube, K. Y. Goh, and P. K. Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43:98–108, February 2006.

[20] S. Nadarajah and S. Kotz. R programs for computing truncated distributions. *Journal of Statistical Software*, 16(2), 2006.

[21] J. Pearl. *Causality: Models, Reasning and Inference*. Cambridge University Press, 2000.

[22] J. Pearl. Direct and indirect effects. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, UAI '01, pages 411–420, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[23] D. B. Rubin. Direct and indirect causal effects via potential outcomes*. *Scandinavian Journal of Statistics*, 31(2):161–170, 2004.

[24] D. B. Rubin. Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[25] D. B. Rubin et al. Causal inference through potential outcomes and principal stratification: application to studies with censoring due to death. *Statistical Science*, 21(3):299–309, 2006.

[26] X. Shao and L. Li. Data-driven multi-touch attribution models. In *Proceedings of the 17th ACM SIGKDD*, pages 258–264, 2011.

[27] P. Wang, Y. Liu, M. Meytlis, H.-Y. Tsao, J. Yang, and P. Huang. An efficient framework for online advertising effectiveness measurement and comparison. In *Proceedings of the 7th ACM WSDM*, pages 163–172, New York, NY, USA, 2014. ACM.

[28] T. Yildiz and S. Narayanan. Star digital: Assessing the effectiveness of display advertising, March 2013. Harvard Business Review: Case Study.

[29] J. L. Zhang and D. B. Rubin. Estimation of causal effects via principal stratification when some outcomes are truncated by death. *Journal of Educational and Behavioral Statistics*, 28(4):353–368, 2003.