



# Digital public health interventions at scale: The impact of social media advertising on beliefs and outcomes related to COVID vaccines

Susan Athey<sup>a,1</sup> , Kristen Grabarz<sup>b</sup>, Michael Luca<sup>b</sup> , and Nils Wernerfelt<sup>c</sup>

Edited by Jose Scheinkman, Columbia University, New York, NY; received May 10, 2022; accepted November 10, 2022

Public health organizations increasingly use social media advertising campaigns in pursuit of public health goals. In this paper, we evaluate the impact of about \$40 million of social media advertisements that were run and experimentally tested on Facebook and Instagram, aimed at increasing COVID-19 vaccination rates in the first year of the vaccine roll-out. The 819 randomized experiments in our sample were run by 174 different public health organizations and collectively reached 2.1 billion individuals in 15 languages. We find that these campaigns are, on average, effective at influencing self-reported beliefs—shifting opinions close to 1% at baseline with a cost per influenced person of about \$3.41. Combining this result with an estimate of the relationship between survey outcomes and vaccination rates derived from observational data yields an estimated cost per additional vaccination of about \$5.68. There is further evidence that campaigns are especially effective at influencing users' knowledge of how to get vaccines. Our results represent, to the best of our knowledge, the largest set of online public health interventions analyzed to date.

public health advertising | meta-analysis | COVID-19 vaccine | randomized experiment

Throughout the rapidly evolving COVID-19 pandemic, policymakers and public health agencies needed to communicate with citizens about mitigation measures ranging from mask wearing and social distancing to vaccines. Advertising on social media emerged as a popular channel to quickly reach large numbers of people and has been used by public health organizations in nearly every country both to convey information and influence behavior. An understanding of the expected impact of these campaigns is important as such organizations continue to engage in interventions as the pandemic unfolds. Assessing these campaigns is further valuable as digital public health interventions become increasingly used to address broader health-related outcomes.

To speak to these questions, this paper aims to evaluate the impact of social media advertisements on a variety of COVID-19-related outcomes. Analyzing advertising campaigns run on Facebook and Instagram by 174 public health organizations around the world, we investigate three main questions. First, what effect did these social media advertising campaigns have? Second, how cost-effective were they? Third, which types of outcomes have the campaigns been most effective at influencing?

The campaigns in our sample were run between December 2020 and November 2021, reached users in nearly every country, and in aggregate consist of \$39.4 million dollars of advertising spending. They include a wide range of public health organizations that span major multinational nonprofits, public health ministries, and local nongovernmental organizations. The identities of the individual advertisers are not included in this article to protect their confidentiality. Importantly, our dataset contains the near universe of relevant experiments that were run on Facebook and Instagram over this period.\* This feature of our sample allows us to draw conclusions that are not vulnerable to selection biases that commonly arise in meta-analyses, such as publication bias, whereby experiments with positive outcomes are more likely to be included in the sample. To our knowledge, the dataset we analyze is the largest set of online public health interventions studied to date.

\*We discuss the nature of these experiments in more detail below and in *SI Appendix*. We are able to identify all campaigns that made use of the platform's standardized "Brand Lift Study" infrastructure as defined later and asked one or more questions about COVID-19 vaccines. The vast majority of experiments identified this way are included in our analysis; the only ones that were excluded were a small number of experiments that chose to ask customized COVID-related questions that do not fall into the seven types that we study. We note that a primary alternative approach for advertisers to conduct randomized experiments is to randomize advertising exposure by geographical region and then compare aggregate outcomes, measured separately, across geographies. We are aware of a handful of such experiments that were run during the time window we consider; these are not included in our analysis because we do not have access to the details of these experiments.

## Significance

This paper analyzes whether social media advertising can be a cost-effective tool to influence attitudes and beliefs about COVID-19 vaccines and ultimately vaccinations. We conduct a meta-analysis of more than 800 experiments conducted by 174 public health-related organizations. Each experiment randomly assigned advertisements to a subset of users and measured the outcome, attitudes, and beliefs, through surveys. Although each experiment individually has too few subjects to reliably detect small effects, by pooling the data, we obtain more precise estimates of the overall average effect. The estimated average cost per additional influenced person, \$3.41, and the estimated cost per additional vaccination, \$5.68, can be compared to the social benefit when evaluating the cost-effectiveness of public health campaigns.

Competing interest statement: The authors have organizational affiliations to disclose. K.G. is a part-time contractor through PRO Unlimited, a contracting agency used by Meta Platforms. N.W. is an employee of Meta Platforms. Yes, the authors have stock ownership to disclose, N.W. owns stock in Meta Platforms. Yes, the authors have research support to disclose, S.A. received funding from Meta Platforms for research projects related to public health. Yes, the authors have additional information to disclose, S.A. previously provided consulting services for Meta Platforms.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: [athey@stanford.edu](mailto:athey@stanford.edu).

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2208110120/-DCSupplemental>.

Published January 26, 2023.

The data have two key features relevant to our analysis. First, we have data from a large number of the campaigns that conducted experiments where exposure to the ads was randomized at the user level, allowing us to assess the causal effect of each campaign. This is especially important in the context of online advertisements, where selection bias is a significant obstacle in nonexperimental data (1–4). Randomized experiments have become more common in online advertising, and companies (including Meta) have developed standardized experimentation tools to facilitate testing. The campaigns we analyze all used these tools to conduct experiments.

Second, we are able to combine the experiments with user-level survey data for a subset of users. The surveys ask a variety of questions, namely, a user's willingness to get a COVID-19 vaccine, belief in the importance of vaccination, belief in vaccine effectiveness, belief in vaccine safety, whether the advertiser is a trustworthy source of COVID-19 information, how knowledgeable the user feels about how to get the vaccines, and whether they think vaccines are socially acceptable (in this paper, we will refer to these questions shorthand as Willingness, Importance, Effectiveness, Safety, Trustworthy Source, Knowledgeable, and Social Norms, respectively). While not all advertisers asked all questions, the survey questions were largely standardized across campaigns, facilitating comparisons. As is common practice with such experiments, we classify responses into a binary outcome according to whether respondents were engaging in the public health behavior of interest (for instance, intending to get the vaccine) or had the relevant public health information (for instance, knowing where to get the vaccine). Looking across all studies, we can then see whether interventions had an impact on the binary outcome of interest (the wording of each question, possible responses, and how answers were classified are all listed in *SI Appendix*). This approach is similar in spirit to the approach taken by other papers that aggregate a set of experiments with distinct outcomes.<sup>†</sup>

Overall, our combined findings suggest that these campaigns were effective at influencing peoples' attitudes and beliefs about the vaccine. We find an average increase in the fraction of positive responses of 0.55% points ( $P = 2e-13$ ) across all experiments, with a baseline 55.7% positive rate. While this point estimate is small on a per-person basis, the reach of the campaigns implies that even under conservative assumptions, around 11.6 million individuals were influenced by these campaigns alone, at a cost of about \$3.41 per incremental person. Translating this estimate into a cost for incremental vaccinations requires additional assumptions and data; the survey outcome can be considered a "surrogate" for the outcome of interest, vaccination (8).<sup>‡</sup> In *SI Appendix*, we present the results of an analysis where we use data from the United States to estimate the correlation between county-level vaccination rates and county-level survey responses, finding a correlation of .6 (standard error .0174) in a sample of 2,710 counties with more than 20 survey responses. Using this estimate

<sup>†</sup>For instance, (5) aggregates a variety of distinct binary outcomes, corresponding to whether an action was taken or not, from a large set of behavioral experiments. Examples include whether or not someone filled out a government form or whether or not someone paid a fine. Two other studies that similarly use meta-analytic methods to combine different treatments and outcomes are refs. 6 and 7.

<sup>‡</sup>One required assumption is that the effect of the treatment is fully captured by the survey question; since the treatment is unlikely to have a negative effect on vaccination, a violation of this assumption would likely lead to a conservative estimate. A second requirement is that the treatment does not directly change the relationship between the survey outcome and vaccination. This requirement could be violated if treatment induced individuals to respond positively to the survey, e.g., in an attempt to please the experimenting organization. This problem is unlikely because the survey is given at a separate time and format than the advertising exposure and is not associated with the public health organization.

together with the result of our meta-analysis implies that the cost of an incremental vaccine is \$5.68. These estimates suggest that campaigns may be an easily scalable intervention that can in aggregate shape the public health outcomes of a large number of citizens.<sup>§</sup> A limitation of this cross-validation approach is that we are ultimately interested in the relationship between marginal survey responses and marginal vaccination rates in response to the ads, whereas we observe a cross-sectional relationship between average survey responses and average vaccination rates at the county level. Our cost calculation approach is based on an assumption that a change in survey responses due to treatment would lead to a change in vaccination rates matching the observed cross-sectional correlation. Since we cannot link treatment to changes in vaccination status, we cannot directly assess whether a change in survey responses due to treatment would lead to a change in vaccination rates that matches the observed cross-sectional correlation. This requirement would fail if, for example, the effect of advertising was very short-lived or if there are barriers to vaccinations among individuals close to the margin of getting vaccinated that are not reflected in the average relationship between survey responses and vaccination decisions.

These results can be broadly compared to those from other initiatives aimed at influencing vaccination decisions.<sup>¶</sup> Barber and West (11) and Sehgal (12) estimate a \$68 and \$49 cost per incremental COVID-19 vaccination in Ohio from the Vax-a-Million lottery (though see refs. 13 and 14); in a separate study, Campos-Mercade (15) find a cost per incremental vaccination on the order of \$400; and Krieger et al. (16) estimate costs of \$88 to \$380 for incremental flu vaccinations in seniors in the United States. Larsen et al. (17) find a much lower cost, specifically \$1 per incremental COVID-19 vaccination from a location-randomized YouTube advertising experiment in select counties in the United States. The World Health Organization (18) has also explored methods of effective communication for influencing health outcomes, including through randomized experiments run on Meta. Since our data do not include information about actual vaccination decisions, it is hard to directly contrast our estimates to those. However, such estimates highlight how challenging it can be to influence health behaviors and the potential value of identifying low-cost, scalable interventions. When considering such interventions to influence vaccine uptake, there has been much advocacy for behaviorally informed promotions (19, 20). The campaigns we analyze broadly fit into this category.

Our findings also connect to the literature on health nudges, which similarly tends to focus on low-cost, scalable interventions. Much of this literature has focused on text-based interventions, which have shown potential across a number of domains, ranging from flu appointments to court appearances (21–23). Specifically, in the context of COVID-19, Dai et al. (24) sent participants in California text-based reminders to make vaccination salient and easy to remember. They find that reminders sent 1 d and 8 d after notification of vaccine eligibility increased vaccination rates by 3.57% points and 1.06% points, respectively. Banerjee et al. (25) analyze the effect of a video

<sup>§</sup>This scalability includes not just reaching more people but also reaching individuals who may be hard or costly to reach via other means. This may be particularly important for some subpopulations where there is evidence of a disproportionate impact of the pandemic (9).

<sup>¶</sup>A separate branch of literature has evaluated impacts of interventions on other COVID-related behaviors. For example, Chen et al. (10) use smartphone data from 10 million devices and find large effects of stay-at-home orders on both movement and transmission rates.

message randomly distributed via SMS to millions of individuals in West Bengal, India; they find substantial effects on both the treated individuals as well as nontreated community members on a broad range of COVID-related outcomes. One advantage of text-based interventions is that they may be more salient and thus have a larger effect relative to our effects. On the other hand, advertising campaigns do not require gathering phone numbers and may thus be more easily scaled to a large population.<sup>#</sup>

Our results can be related to prior large-scale meta-analyses of online advertising. The literature has highlighted major challenges due to low statistical power (28). Meta-analysis is a natural way to address this challenge, but it requires access to data from the experiments of many advertisers and also creates challenges comparing effectiveness across heterogeneous advertising objectives. We are aware of only three other meta-analyses of digital advertisements that have a comparable scale to our study.<sup>||</sup> First, Johnson et al. (29) used internal data from Google's display advertising platform to study the effect of digital advertisements on website visits for 432 digital advertising campaigns, finding effects of 8% of baseline website visits. Second, Goldfarb and Tucker (32) analyzed 2,892 experiments carried out by a brand research firm, each using a similar survey methodology and sample size to the experiments considered here, finding an effect of about 10% of the baseline on survey responses concerning intention to purchase. Third, Gordon et al. (33) analyze more than 600 advertising experiments on Facebook, comparing the estimated effects for different measures along what is referred to as a "funnel" or a customer's journey to a final action of interest. They find effects of 28%, 19%, and 6% for measured outcomes that capture consumer behavior at the top, middle, and lower parts of the funnel. In our paper, we conduct a large-scale meta-analysis of online public health campaigns across multiple outcomes. Our estimated effects (about 1% over baseline) are substantially smaller than the effects found in these studies, suggesting that it is more difficult to change attitudes and beliefs about vaccination than it is to increase more standard advertising outcomes. Similar to ref. 33, we find smaller effects for outcome measures that are closer to the ultimate outcome of interest, vaccination.

Building on our main results, we next look across the different survey outcomes to see which ones are most impacted by the campaigns. We find significant effects on Knowledge, Safety, Social Norms, and Importance (all have  $P < 0.001$ ), while no significant effects on Willingness, Effectiveness, or Trustworthy Source. There were individual campaigns that were able to significantly move these last three metrics, but we could not detect an overall average effect. Finally, we find evidence that the campaigns may have been particularly effective at shifting users' knowledge around the vaccines. Knowledge has the largest treatment effect point estimate (1.23% points,  $P = 5e-7$ ), and it is significantly higher than nearly all the other coefficients.\*\*

<sup>#</sup>A related vein of literature has focused on identifying mechanisms for effective communication around COVID-19 that could then be implemented at scale. For example, holding a wide range of factors constant, Alsan and Eichmeyer (26) vary characteristics of the messenger and signal content in a video infomercial and find evidence of substantial heterogeneity in effectiveness only by shifting those attributes. Similarly, Jordan et al. (27) find evidence that prosocial framings are important for shifting COVID-related outcomes across a range of experiments. Insights from studies such as these two could help inform both digital and nondigital interventions.

<sup>||</sup>Ref. 29 provides an overview of the literature, including related meta-analyses in other forms of advertising such as online search advertising and television; see also ref. 30 that analyzed 54 mobile advertising campaigns and ref. 31 that studied social ads across 74 products.

\*\*In a two-sided  $t$ -test, the coefficient for Knowledge is significantly greater than that for Effectiveness ( $P = 0.038$ ), Importance ( $P = 0.003$ ), Safety ( $P = 0.034$ ), Trustworthy Source

This suggests that, on average, the digital advertisements in our sample may have been a particularly cost-effective channel for information dissemination or that information is easier to retain. In interpreting these comparisons, it is important to recall that different campaigns were designed for different purposes, and some campaigns may have conducted experiments on both primary and secondary outcomes so that smaller impacts might be expected for secondary outcomes. For example, a campaign focused on providing information might have evaluated its impact on Knowledge but also on Willingness. If Knowledge was less commonly included as a secondary outcome, it might be expected that the measured impact would be higher. In *SI Appendix*, we investigate this hypothesis by manually classifying advertisements as to which outcomes are their primary ones, and we find similar patterns (our manually labeled sample is smaller, but we see that the point estimate for Knowledge is again significantly larger than that for Importance, Safety, and Willingness; it is nonsignificantly different from that for Effectiveness and Social Norms).

Overall, our results suggest that social media advertising campaigns can be an important component of public health initiatives. Over the course of the past 2 y, health-oriented organizations have engaged in a wide range of tactics in an effort to shift attitudes and behaviors (see <https://www.nga.org/center/publications/covid-19-vaccine-incentives/> for a list of different incentives offered in the United States alone, ranging from Girl Scout cookies to laps on a NASCAR track); key challenges with many of these include scalability, measurement, and generalizability. Digital advertising can help overcome these challenges. However, the small per-person impact highlights that these campaigns are best thought of as part of a broader set of strategies. To this end, our paper complements the growing literature on designing effective public health interventions for COVID-19.

The rest of our paper is organized as follows. The *Data* describes our sample, the outcome variables, and the approach we use to analyze the campaigns; the *Results* presents findings in greater detail; and the *Conclusion* summarizes. Additional analyses are provided in *SI Appendix*.

## Data

**Overview.** We analyze a set of 819 randomized experiments that were conducted between December, 2020, and November, 2021. The experiments in our sample are derived from 376 distinct advertising campaigns and 174 organizations. The advertising experiments and surveys were conducted by Meta and the public health agencies prior to our analysis. There are often multiple experiments associated within a single ad campaign, where each experiment corresponds to a specific survey outcome. For example, an advertiser may take one campaign and run three separate experiments that measure the impact of the campaign on Willingness, Importance, and Effectiveness. The average campaign in our data ran slightly more than two experiments; in other words, advertisers measured the impact of their campaigns on an average of about two outcome metrics each (the platform normally caps the number of questions per campaign at three).

The studies were all conducted using Meta's infrastructure for conducting advertising effectiveness experiments across Facebook and Instagram; through the rest of this paper, we refer to this infrastructure as "the platform." We focus specifically on experiments that measure the extent to which advertisements

( $P = 0.002$ ), Willingness ( $P = 0.020$ ), and our overall estimate ( $P = 0.008$ ), including when dropping Knowledge studies ( $P = 0.006$ ). It is not significantly different from our estimate for Social Norms, though it is close in a one-sided test ( $P = 0.109$ ).



**Table 1. Summary statistics by outcome metric**

Category	Effectiveness	Importance	Knowledge	Safety	Social norms	Trustworthy source	Willingness	Overall
# Experiments	64	237	94	218	107	59	40	819
# Unique campaigns	48	234	73	218	107	59	40	376
# Unique organizations	32	109	57	100	50	30	17	174
Earliest campaign start	2021-02-17	2020-12-08	2021-02-19	2020-12-15	2021-02-03	2021-02-24	2020-12-08	2020-12-08
Latest campaign end	2021-11-15	2021-11-12	2021-11-03	2021-11-15	2021-11-14	2021-10-21	2021-06-15	2021-11-15
Avg. # people reached	17,932,855 (3,216,101)	12,833,435 (1,353,134)	15,341,751 (2,957,774)	11,959,068 (1,746,021)	12,497,147 (2,396,430)	9,510,847 (1,971,257)	27,297,621 (6,773,069)	12,913,047 (1,225,276)
Avg. # survey resp per experiment	1,753.53 (151.09)	2,130.65 (90.27)	1,734.60 (128.96)	1,931.67 (90.81)	1,510.58 (120.24)	1,762.17 (148.21)	1,429.00 (155.72)	1,860.94 (45.75)
Avg. campaign cost	\$136,807 (\$41,543)	\$65,487 (\$6,376)	\$122,796 (\$36,000)	\$132,491 (\$26,694)	\$63,556 (\$13,409)	\$104,177 (\$24,353)	\$392,678 (\$142,627)	\$105,183 (\$16,872)
# Experiments rejecting no effect (0.1)	7	23	22	29	14	3	1	99
# Experiments rejecting no effect (0.05)	5	16	17	20	8	3	1	70
Implied false discovery rate (0.1)	0.914	1.000	0.427	0.752	0.764	1.000	1.000	0.827
Implied false discovery rate (.05)	0.640	0.741	0.276	0.545	0.669	0.983	1.000	0.585
# FDR survivor experiments (10% FDR)	2	14	5	5	1	2	0	27

Note: Standard errors in parentheses. # Experiments Rejecting No Effect (0.1) references the number of studies that are significant at the 0.1 level in a two-tailed *t*-test against the null of no treatment effect. Implied False Discovery Rate (0.1) estimates the false discovery rate if we accepted all experiments that were significant at the 0.1 level, as  $FDR(0.1) = (0.1 * n_e) / (n_{rej})$ , where 0.1 is the level of significance,  $n_e$  is the number of experiments, and  $n_{rej}$  is the number of experiments rejecting no effect at 0.1. # FDR Survivor Experiments is the number of experiments determined via the Benjamini–Hochberg algorithm to survive a false discovery rate of 10% (34).

affect individuals’ attitude or beliefs as measured by survey questions.<sup>††</sup>

We limit the set of experiments to those measuring outcomes in one of seven categories mentioned earlier (*SI Appendix* for details).<sup>‡‡</sup> Though not an exhaustive set of COVID-19 ad experiments, these seven categories were selected because they are the most prevalent across COVID-19 vaccine-related experiments. We use all studies that asked these questions and, following the platform’s policy (see <https://www.facebook.com/business/help/2396060560411130>), restrict to users aged 18 y and older. For our analysis, we used deidentified data for all campaigns.

The campaigns we study total \$39.4 million in ad spend, with a reach of 2.1 billion unique users translated across 15 languages. The average campaign costs more than \$100,000 and reached nearly 13 million people; these were substantive efforts, but importantly also not beyond the budget of what many public health organizations could conceivably spend on similar campaigns in the future. Table 1 provides summary statistics.

Although the campaigns we study reached billions of users, as we can see in Table 1, we observe a much smaller number of survey responses. The platform provides the experimentation service to enable advertisers to estimate the incremental impact of their campaigns on survey-based outcomes; however, the number of responses per experiment is limited by the platform.

<sup>††</sup> These experiments are known as “Brand Lift Studies” in advertiser-facing documentation. There are many companies that offer Brand Lift experiments to advertisers, each with slightly different implementations and methodologies. These studies are commonly used to measure effects on outcomes such as ad recall, brand sentiment, or intent to purchase but have become popular during the pandemic to also look at health-related outcomes that may not be observable in log data). We note that not all advertisers run these experiments, so our results are underestimates of the total impact of digital advertising interventions on Facebook and Instagram (Meta imposes minimum budgets to run one of these studies that vary across country; for example, in the United States, it is currently \$30,000, which is more than many advertisers’ budgets).

<sup>‡‡</sup> While the platform proposed standardized questions to the advertisers, they did have autonomy to adjust the language for the questions if they wished, so that we see some heterogeneity of questions asked within the seven categories we study. In our data, there are no instances where the same campaign ran multiple experiments that each asked the exact same survey question. However, 38 campaigns ran multiple experiments measuring the same outcome variable using distinct questions. For example, a campaign may have run two separate experiments that both measured its impact on Knowledge, with different survey questions, such as Do you know your order of priority to get the COVID-19 vaccine? and Do you know where to go to get a COVID-19 vaccine for yourself? Omitting these campaigns yields no material shift in our results, and in *SI Appendix*, we conduct additional analyses that factor in this within-campaign heterogeneity.

The limits are presumably motivated by the fact that users may be willing to engage in only a small number of surveys, and the user experience may be negatively impacted by too many surveys. Thus, the platform caps the number of respondents per study to balance the tradeoff between statistical power and user experience (*SI Appendix* for more details on how these experiments are implemented).

In total, our dataset incorporates 1.5 million total responses across all experiments. Per experiment, the number of responses ranges from 300 to 4,507, split across test and control groups. This does not give us too much power for each individual experiment. If an experiment of average sample size (1861 in our sample) and average baseline positive response rate (55%) were to be analyzed using a difference in means between treatment and control, the minimum detectable effect size with 80% power and a 10% significance level would be about 0.06, close to 10% of the baseline. The average effects we find below are an order of magnitude lower than that (in response to an early draft of this manuscript, the platform started granting more exceptions to the normal response caps on these experiments in order to deliver better-powered results). Indeed, we see that only 99 out of the 819 experiments rejected the null at the 10% level, just slightly greater than the 10% of experiments that would be expected if there were no treatment effects.

Only 27 experiments on their own are included in a set of experiments determined to have a 10% false discovery rate using the Benjamini–Hochberg procedure (34). We also report the implied false discovery rate for the set of experiments that are individually significant at the 5% level, and we repeat this for the 10% level. This exercise is motivated by the idea that an organization might choose to further scale a campaign after seeing a statistically significant impact. We see that if organizations used the 5% threshold for scaling, for the Knowledge outcome, the false discovery rate would be about 1/4, while for the Importance outcome, the rate would be about 3/4. Although such a false discovery rate might not be problematic, as it is unlikely that the campaigns would be harmful and in aggregate these campaigns would be cost-effective, our findings as a whole suggest that we are not well powered to well identify individually effective campaigns. This motivates the approach we pursue in this paper of conducting a meta-analysis of hundreds of experiments

**Table 2. Overview of survey questions**

Question category	Wording
Importance	How important do you feel a vaccine is to prevent the spread of COVID-19?
Safety	How safe do you think a COVID-19 vaccine is for people like you?
Willingness	How likely are you to get vaccinated for COVID-19 when the vaccine is available to you?
Effectiveness	How effective do you think the COVID-19 vaccination is in preventing COVID-19?
Knowledge	Do you know where people in your local community can go to get a COVID-19 vaccine?
Social norms	When you think of most people whose opinion you value, how much would they approve of people getting a COVID-19 vaccine?
Trustworthy source	Do you agree or disagree that [advertiser name] is a trustworthy source of COVID-19 vaccine facts and information?

together rather than seeking to identify individual campaigns with positive effects.

In *SI Appendix*, we provide graphs of the CDFs of  $P$ -values both overall and for each metric, where we can see that overall and particularly for the Knowledge outcome, the CDFs of  $P$ -values depart from the uniform distribution that would be expected if there were no effects.

**Survey Questions.** We provide more detail on the questions, their possible responses, and their coding in *SI Appendix*. In Table 2 we provide text from our standardized survey questions as a reference.

**Validity of Survey-Based Outcomes.** We now turn to consider the validity of the survey-based outcome measures, and in particular, the extent to which they (or changes in them) do not capture changes in beliefs or knowledge. Privacy and legal constraints prevent advertisers from asking about or measuring some ultimate quantities of interest on platform (e.g., health or vaccination status), but the self-reported measures may still be meaningful.<sup>§§</sup> Here, we discuss two categories of potential concerns about these measures.

A first category of concerns relates to whether the survey outcomes as entered in the platform reflect beliefs and behaviors in the physical world. There are several considerations. Following established practice for social media brand campaigns, survey outcomes are a primary outcome that public health organizations have been using to evaluate their campaigns. Campaigns start, stop, and change based on the results of these experiments, dictating how entire ad budgets of COVID-19 interventions are spent. Hence, it is important to understand how these outcomes have responded to campaigns to date and to add to the understanding of public health organizations whose individual experiments to date have been underpowered to detect small effects.

Relatedly, a common goal of public health organizations is to simply shift attitudes and beliefs. Akin to traditional advertisers where campaigns may target different levels of the conversion funnel, many of these advertisers are aiming to move awareness or basic beliefs and may invest in complementary tactics to change behavior once beliefs have been influenced. To the extent that the implementation details and survey responses provide insight into awareness and beliefs, the campaign experiment outcomes are informative.

Finally, for social media campaigns in general, there is evidence that responses to platform surveys correlate reasonably well

with behaviors of interest. Moehring et al. (36) find an  $R^2$  of 0.83 in a regression of country-level vaccine uptake on self-reported vaccine status collected from a survey on Facebook. And Astley et al. (37) find a correlation between survey metrics on Facebook and off-platform COVID-19 cases (see refs. 38 and 39 for further discussions and caveats). Alekseev et al. (40) find a high degree of correlation between characteristics of businesses that Facebook users self-report to own and offline statistics from the US Census. Although the contexts and analyses from these studies are different, together, they suggest that there does appear to be informative signal in social media survey outcomes.

In *SI Appendix*, we explore the relationship between survey response positivity and county-level vaccine takeup and find that the two are strongly correlated. We use these findings to extrapolate an estimated cost of each additional vaccination.

A second category of concerns about survey results relates to whether the differences in survey outcomes between treated and control groups can be interpreted as the causal effect of the advertisements (1). One issue is that there can be systematic differences between the treatment and the control group due to the implementation of the randomized advertising experiment. Details are provided in *SI Appendix*, but in short, randomization of assignment to ads takes place just before an ad was intended to be shown to a user, so that whether a user is sent an ad is random within the experiment. However, after seeing an ad, whether a user in the treatment group is subsequently shown a survey depends on an additional factor that is not present for the control group. In particular, if after randomization into the treatment group, the platform intends to show the user an ad but the user scrolls past it or does not scroll to it at all, the user will not be sent a survey. This is done to only capture survey responses from users in the treatment group who actually saw the ad, but to the extent to which this behavior is correlated with the survey outcome, it could lead to confounding and thus biased estimates of treatment effects.

In addition, even if the set of users who were sent surveys was perfectly randomized within each experiment, there still is the potential for differential survey response between the treatment and the control group. This might occur if individuals influenced by the ads were more likely to respond to the survey. We address both of these issues in *SI Appendix*, where we show that along several observable dimensions, the treatment and control groups are similar. We further address these concerns by adjusting for several observable characteristics of individual respondents in our analysis.

A final concern is that the population answering the surveys differs from the overall target population along unobservables correlated with our outcome variable. For example, certain age groups may be more likely to answer the surveys. Poststratifying our results by age and gender (as we do) is an industry

<sup>§§</sup>Breza et al. (35) ran a location-randomized experiment where different regions were targeted with ad campaigns on Facebook. This experimental design allowed measurement and detection of significant effects on off-platform outcomes namely, travel and actual COVID-19 cases. This result, though from a single experiment, demonstrates the potential for relevant offline effects from similar digital ads.

standard approach to address this concern; in addition, in *SI Appendix*, we also compare observables from a poststratified sample with those from the target population and find reasonable overlap.

## Methods

We conduct a meta-analysis by first analyzing each experiment separately and then using inverse variance weighting to generate the average effect for each outcome and overall.

**Meta-Analysis of Experiments.** We begin by analyzing each experiment separately using the following weighted linear model:

$$\text{response}_i = (\tilde{X}'_i \beta + \beta_0) W_i + \tilde{X}'_i \gamma + \varepsilon_i, \quad [1]$$

where  $\text{response}_i$  is an indicator for whether individual  $i$  gave a positive response,  $\tilde{X}_i$  is a matrix of de-meaned controls that could potentially be related to outcomes (age bucket, gender, expected click-through rate, and expected conversion rate), and  $W_i$  is a dummy variable denoting whether  $i$  was in the treatment or control group. The expected click-through and conversion rates are platform-generated estimates age buckets are 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65+. We de-mean and interact our covariates with the treatment indicator so that  $\hat{\beta}_0$  remains an unbiased and consistent estimate of the average treatment effect even in the presence of heterogeneous treatment effects by our covariates (41).

In our regression, each response is weighted with poststratification weights by age bucket and gender within the treatment and control group, such that both arms of each experiment are representative of the population reached by the relevant campaign. That is, we obtain the proportion of users in each age bracket and gender group reached by the campaign associated with a given experiment and divide it by the proportion of responses to obtain the weight for each response (to reduce variance, they are trimmed to an upper bound of 3 and a lower bound of 0.3 to decrease the influence of outlying observations. Rerunning without any trimming yields no material difference in the results). We poststratify by these two variables as age and gender are basic demographics that advertisers are both frequently interested in and where heterogeneous effects are often observed.<sup>¶¶</sup> In *SI Appendix*, we explore robustness to different weighting schema and find no material difference in the results.

This procedure is based on the one that Meta uses to analyze results from these survey-based experiments for advertisers. Repeating this approach for each of the experiments in our dataset, we are left with 819 estimates of average treatment effects and standard errors. The next step in the analysis is to combine these point estimates into estimated effects by outcome metric and to generate an overall, combined estimate, following standard meta-analytic methods (42, 43).

Specifically, to generate the average effect for each outcome and overall, we combine the respective experiments using inverse variance weighting. This approach estimates a single, homogeneous effect per category while minimizing variance. We present this approach due to its simplicity and the fact that there is no evidence of heterogeneity across all outcomes; in *SI Appendix*, we report several alternative specifications that allow for greater heterogeneity across experiments and outcomes and find very similar results.

<sup>¶¶</sup> At the experiment level, we observe many significant positive and negative coefficients on our age and gender interaction terms; rerunning our meta-analyses on these coefficients yields insignificant average effects, however.

## Results

We now turn to our main results, which we break into two categories. First, we describe for each survey outcome, the average effect across all experiments that focused on that outcome. Second, we combine the treatment effect estimates with data about the number of unique people who received advertisements as well as the total cost of the campaigns to estimate how many people have been influenced and what the cost per influenced person is.

The results of our meta-analysis, including the average effect for each outcome and overall, are described in Table 3.

Several comments are salient to the interpretation of these results. First, this dataset is very broad. Past efforts to understand what has and has not worked with shifting behaviors around COVID-19 have often by necessity studied a small number of treatments at modest scale or been one off ex post analyses. External validity with such studies is frequently a concern and potentially helps explain why studies to date have found conflicting results (e.g., refs. 12–15, 44 on effects of financial incentives on vaccination rates). In contrast, here, pooling hundreds of studies from a broadly representative population, we find a positive and statistically significant average main effect. The evidence may not be conclusive yet on what kinds of behavioral nudges work, but this is evidence that these digital advertising campaigns can help move the needle on COVID-related attitudes.

Second, consider results about the specific outcomes. We find that Importance, Knowledge, Safety, and Social Norms showed highly statistically significant effects. In contrast, we do not detect an effect for Effectiveness, Trustworthy Source, or Willingness. (Though Effectiveness is close to marginally significant in our main specification.)

We note that these last three metrics, particularly Effectiveness and Willingness, are arguably lower in the vaccine conversion funnel (that is, they are better proxies for a final desired action) than the first four metrics. A stylized fact from advertising is that such “lower-funnel” behaviors often see smaller effect sizes than more upper-level outcomes and are generally challenging to study as they may be influenced by a variety of unmeasured factors. While our finding thus accords with this intuition, a limitation of our study is that even with a very large sample size, we are not powered to generate more precise estimates of these averages (see the final rows of Table 3).

Third, we note that there is evidence that the estimated lift for Knowledge (1.23pp) is significantly higher than our other metrics. Our estimate for Knowledge is significantly greater than that for Effectiveness ( $P = 0.038$ ), Importance ( $P = 0.003$ ), Safety ( $P = 0.034$ ), Trustworthy Source ( $P = 0.002$ ), Willingness ( $P = 0.020$ ), and our overall estimate ( $P = 0.008$ ), including when dropping the Knowledge experiments ( $P = 0.006$ ). It is nearly significantly greater than Social Norms in a one-sided  $t$ -test ( $P = 0.109$ ). We note that Knowledge is a distinct outcome here in that the other outcomes relate more to persuasion, whereas Knowledge focuses simply on conveying information. These results suggest that social media campaigns may be particularly attractive for public health organizations interested in the latter.<sup>##</sup> (See also our additional analysis in *SI Appendix* with manually labeled ads).

<sup>##</sup> In thinking about the effects for Knowledge as well as the other metrics, a relevant data point is the baseline positive response rate for Trustworthy Source is 36.5%. While this question was asked only in a subset of campaigns, it is still revealing that the effects we are seeing are despite a relatively low user trust level with some advertisers. This suggests that health organizations with strong brand values may be well positioned to see particularly large effects, a topic we leave to future research.



**Table 3. Meta-analysis of experiments by outcome**

Category	Effectiveness	Importance	Knowledge	Safety	Social norms	Trustworthy source	Willingness	Overall
Treatment coefficient	0.0045 (0.0029)	0.0043*** (0.0012)	0.0123*** (0.0025)	0.0062*** (0.0016)	0.0081*** (0.0024)	0.0012 (0.0027)	0.0010 (0.0042)	0.0055*** (0.0008)
P-value	0.114	0.0004	5e-7	8e-5	0.0006	0.639	0.807	2e-13
Cost per influenced person	\$2.43	\$2.41	\$0.77	\$3.20	\$1.00	\$11.69	\$17.14	\$3.41
Baseline positive response rate	0.505	0.672	0.575	0.501	0.556	0.365	0.517	0.557
Treatment effect as % of baseline	0.89%	0.64%	2.14%	1.24%	1.46%	0.33%	0.19%	0.99%
Power calculations (approximate)								
Minimum detectable effect	0.0071	0.0030	0.0061	0.0039	0.0059	0.0066	0.0104	0.0019
Power to detect given effect size	0.474	0.973	1.000	0.989	0.962	0.120	0.081	1.00
# experiments needed for 80% power	159	115	23	87	57	1,660	4,133	94
# survey resp. per exp for 80% power	4,349	1,036	426	770	801	49,567	147,635	214

Note: Standard errors in parentheses. For each column, we consider the set of associated experiments and calculate the inverse-variance weighted average treatment effect (row 1). The baseline % positive response is an unweighted mean across all the relevant experiments; calculating it using fixed or random effects models changes the numbers only slightly. The cost per influenced person for each subset is calculated using the spend and number of unique people reached across all campaigns in the relevant subset. Finally, we include power calculations based on the standard error of the treatment effects (abstracting away from heterogeneity across experiments). Power to detect a given effect size is calculated at the  $\alpha = .1$  level; for the last two rows, we want to convey how power could be improved by either increasing the number of experiments or the number of surveys per experiment. For those calculations, we ask how much of either we would need holding the other fixed to have 80% power to detect the given estimated treatment effect. *SI Appendix* for details.

Finally, we note that in *SI Appendix*, we explore different specifications and find broadly similar results. In addition, the standard output provided to advertisers on Meta comes from a hierarchical Bayesian model; we chose a frequentist approach due to its simplicity, but in *SI Appendix*, we show robustness to using a similar Bayesian approach.

#### Number of Influenced People, Cost per Influenced Person.

Conditional on our results, how many people were influenced by these campaigns, and how cost-effective were they? As aforementioned, the survey data comes from only a subset of the overall users who saw the ads; to calculate the number of influenced people, we follow the common industry practice and scale the point estimate of the treatment effect from each experiment by the size of the overall population that saw the campaign. In our case, since we have data across many advertisers and some users were shown ads from multiple campaigns, to generate a (conservative) estimate of the number of influenced people per campaign, we can treat these collective campaigns as effectively one large one. Specifically, we combine the total spend, total unique reach, and our estimate of the average treatment effect to generate a back-of-the-envelope estimate of the number of people who were influenced by this combined effort.

Doing this calculation, we estimate that about 11.6 million people were influenced by these campaigns. To be clear, by “influenced” we mean shifted self-reported beliefs to a positive outcome; this does not capture people who, for example, moved along the intensive margin of these categories (in *SI Appendix*, we explore a less conservative way of estimating the number of influenced people and thus the cost per influenced person).

Conditional on estimates of how many people were influenced, how much does it cost to influence someone? For this, we divide the number of influenced people by the ad spend, again as is typical in the industry. From Table 3, we can see that the average cost per influenced person was \$3.41.

To understand how cost-effectiveness translates to real-world public health outcomes, in *SI Appendix*, we explore the relationship between survey positivity rate and vaccine series completion rate at the county level in the United States. Across survey outcomes, we find that for each additional positive

survey response, we would expect to see about a 0.6 increase in vaccination takeup in the CDC data. Applying this to our average cost per influenced person from Table 3, this implies an estimated cost per additional vaccine of about \$5.68.

While we are hesitant to extrapolate substantially outside our sample, these magnitudes suggest that running even a few million dollars’ worth of additional campaigns could achieve relatively large shifts in the baseline fraction of outcome variables.

#### Conclusion

Over the course of the pandemic, public health agencies increasingly leveraged social media advertising to pursue public health goals. Our results show that public health interventions via digital advertising are an effective medium for changing important self-reported beliefs and attitudes around COVID-19. Combining with nonexperimental data on vaccination rates, our results suggest that these campaigns were a cost-effective approach to increasing rates as well. The cost-effectiveness and scale of these campaigns can make them appealing to a broad range of organizations around the world. The use of social media advertising more broadly has the potential to aid in the pursuit of other health policy goals, ranging from childhood vaccination to hand washing.

**Data, Materials, and Software Availability.** We are not sharing individual level data as part of this work. However, on <https://www.github.com/gsbDBI/CovidAdMeta/>, we provide a simulated dataset and code to reproduce all our main results.

**ACKNOWLEDGMENTS.** We thank Kang-Xing Jin and the health group at Meta Platforms for leading the effort at Meta surrounding these public health interventions. S.A. acknowledges funding from the Golub Capital Social Impact Lab at Stanford Graduate School of Business. Funding to hire K.G. as a part-time contractor to work on this project came from Meta Platforms.

Author affiliations: <sup>a</sup>Stanford Graduate School of Business, Stanford, CA 94305; <sup>b</sup>Harvard Business School, Boston, MA 02163; and <sup>c</sup>Meta Platforms, Menlo Park, CA 94025

Author contributions: S.A., K.G., M.L., and N.W. designed research; S.A., K.G., M.L., and N.W. performed research; K.G. and N.W. analyzed data; and S.A., K.G., M.L., and N.W. wrote the paper.

1. B. R. Gordon, F. Zettelmeyer, N. Bhargava, D. Chapsky, A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Market. Sci.* **38**, 193–225 (2019).
2. D. Eckles, B. R. Gordon, G. A. Johnson, Field studies of psychologically targeted ads face threats to internal validity. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5254–E5255 (2018).
3. T. Blake, C. Nosko, S. Tadelis, Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica* **83**, 155–174 (2015).

4. S. Moshary, B. T. Shapiro, J. Song, How and when to use the political cycle to identify advertising effects. *Market. Sci.* **40**, 283–304 (2021).
5. S. DellaVigna, E. Linos, RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica* **90**, 81–116 (2022).
6. S. Benartzi *et al.*, Should governments invest more in nudging? *Psychol. Sci.* **28**, 1041–1055 (2017).

7. D. Hummel, A. Maedche, How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *J. Behav. Exp. Econ.* **80**, 47–58 (2019).
8. S. Athey, R. Chetty, G. W. Imbens, H. Kang, "The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely" (Tech. rep., National Bureau of Economic Research, 2019). <https://www.nber.org/papers/26463>. Accessed 22 October 2022.
9. M. Alsan, A. Chandra, K. Simon, The great unequalizer: Initial health effects of COVID-19 in the United States. *J. Econ. Perspect.* **35**, 25–46 (2021).
10. M. K. Chen, Y. Zhuo, M. de la Fuente, R. Rohla, E. F. Long, "Causal estimation of stay-at-home orders on SARS-COV-2 transmission" Tech. rep. arXiv [Preprint] (2020). <https://www.arxiv.org/abs/2005.05469>. Accessed 22 October 2022.
11. A. Barber, J. West, Conditional cash lotteries increase COVID-19 vaccination rates. *J. Health Econ.* **81**, 102578 (2021).
12. N. K. Sehgal, Impact of Vax-a-million lottery on COVID-19 vaccination rates in Ohio. *Am. J. Med.* **134**, 1424–1426 (2021).
13. A. J. Walkley, A. Law, N. A. Bosch, Lottery-based incentive in Ohio and COVID-19 vaccination rates. *JAMA* **326**, 766–767 (2021).
14. H. Thirumurthy, K. L. Milkman, K. G. Volpp, A. M. Buttenheim, D. G. Pope, Association between statewide financial incentive programs and COVID-19 vaccination rates. *PLoS One* **17**, e0263425 (2022).
15. P. Campos-Mercade *et al.*, Monetary incentives increase COVID-19 vaccinations. *Science* **374**, 879–882 (2021).
16. J. W. Krieger, J. S. Castorina, M. L. Walls, M. R. Weaver, S. Ciske, Increasing influenza and pneumococcal immunization rates: A randomized controlled study of a senior center-based intervention. *Am. J. Prevent. Med.* **18**, 123–131 (2000).
17. B. Larsen *et al.*, "Counter-stereotypical messaging and partisan cues: Moving the needle on vaccines in a polarized U.S." (Tech. rep., National Bureau of Economic Research [Preprint], 2022). <https://www.nber.org/papers/w29896>. Accessed 22 October 2022.
18. W. H. Organization, Communication for health in the who western pacific region (2021).
19. J. J. V. Bavel *et al.*, Using social and behavioural science to support COVID-19 pandemic response. *Nat. Hum. Behav.* **4**, 460–471 (2020).
20. K. G. Volpp, G. Loewenstein, A. M. Buttenheim, Behaviorally informed strategies for a national COVID-19 vaccine promotion program. *JAMA* **352**, 125–126 (2021).
21. K. L. Milkman *et al.*, A megastudy of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2101165118 (2021).
22. A. Buttenheim *et al.*, Effects of ownership text message wording and reminders on receipt of an influenza vaccination: A randomized clinical trial. *JAMA Netw. Open* **5**, e2143388 (2022).
23. A. Fishbane, A. Ouss, A. K. Shah, Behavioral nudges reduce failure to appear for court. *Science* **370**, eabb6591 (2020).
24. H. Dai *et al.*, Behavioural nudges increase COVID-19 vaccinations. *Nature* **597**, 404–409 (2021).
25. A. Banerjee *et al.*, "Messages on COVID-19 prevention in India increased symptoms reporting and adherence to preventive behaviors among 25 million recipients with similar effects on non-recipient members of their communities" (Tech. rep., National Bureau of Economic Research [Preprint], 2020). <https://www.nber.org/papers/27496>. Accessed 22 October 2022.
26. M. Alsan, S. Eichmeyer, Experimental evidence on the effectiveness of non-experts for improving vaccine demand (National Bureau of Economic Research [Preprint], 2021). <https://www.nber.org/papers/28593>. Accessed 22 October 2022.
27. J. J. Jordan, E. Yoeli, D. G. Rand, Don't get it or don't spread it: Comparing self-interested versus prosocial motivations for COVID-19 prevention behaviors. *Sci. Rep.* **11**, 1–17 (2021).
28. R. A. Lewis, J. M. Rao, The unfavorable economics of measuring the returns to advertising. *Q. J. Econ.* **130**, 1941–1973 (2015).
29. G. Johnson, R. A. Lewis, E. Nubbemeyer, The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. Available at SSRN 2701578 (2017).
30. Y. Bart, A. T. Stephen, M. Sarvary, Which products are best suited to mobile advertising? A field study of mobile display advertising effects on consumer attitudes and intentions. *J. Market. Res.* **51**, 270–285 (2014).
31. S. Huang, S. Aral, Y. J. Hu, E. Brynjolfsson, Social advertising effectiveness across products: A large-scale field experiment. *Market. Sci.* **39**, 1142–1165 (2020).
32. A. Goldfarb, C. E. Tucker, Privacy regulation and online advertising. *Manage. Sci.* **57**, 57–71 (2011).
33. B. R. Gordon, R. Moakler, F. Zettelmeyer, Close enough? A large-scale exploration of non-experimental approaches to advertising measurement. arXiv [Preprint] (2022).
34. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
35. E. Breza *et al.*, Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: A cluster randomized controlled trial. *Nat. Med.* **27**, 1622–1628 (2021).
36. A. Moehring *et al.*, Surfacing norms to increase vaccine acceptance. Available at SSRN 3782082 (2021).
37. C. M. Astley *et al.*, Global monitoring of the impact of COVID-19 pandemic through online surveys sampled from the Facebook user base. medRxiv (2021).
38. V. C. Bradley *et al.*, Unrepresentative big surveys significantly overestimated us vaccine uptake. *Nature* **600**, 695–700 (2021).
39. A. Reinhardt, R. Tibshirani, "Big data, big problems: Responding to 'are we there yet?'" Tech. rep. (2021). <https://www.arxiv.org/abs/2109.00680>. Accessed 22 October 2022.
40. G. Alekseev *et al.*, "The effects of COVID-19 on us small businesses: Evidence from owners, managers, and employees" (Tech. rep., National Bureau of Economic Research, 2020).
41. G. W. Imbens, D. B. Rubin, *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge University Press, 2015).
42. M. Borenstein, L. V. Hedges, J. P. Higgins, H. R. Rothstein, *Introduction to Meta-Analysis* (John Wiley & Sons, 2021).
43. L. V. Hedges, I. Olkin, *Statistical Methods for Meta-Analysis* (Academic Press, 2014).
44. T. Chang, M. Jacobson, M. Shah, R. Pramanik, S. B. Shah, "Financial incentives and other nudges do not increase COVID-19 vaccinations among the vaccine hesitant" (Tech. rep., National Bureau of Economic Research [Preprint], 2021). <https://www.nber.org/papers/29403>. Accessed 22 October 2022.