Can scientific information address vaccine hesitancy? A randomized controlled trial [*]

Daniel W. Sacks[†] Brigham Walker[‡]

July 13, 2026

Abstract

In recent years pediatric and seasonal vaccination rates have fallen steadily, despite a strong scientific consensus that vaccinations are important for health. We investigate whether communicating scientific information about the flu vaccine can address vaccine hesitancy. We focus on side effect information because prior work has shown that beliefs about side effects are both influential for vaccine decisions and, for vaccine hesitant people, badly overstate risks documented in clinical trials. We find that side effect information from scientific studies can shift beliefs about side effect risks and intentions to vaccinate among vaccine hesitant participants, despite their lower levels of trust in academic and government institutions.

JEL codes: D83, I12, L65
Key words: vaccines, belief elicitation, information provision

[*]We thank seminar audiences at the University of Wisconsin-Madison and the Booth Health Economics Summit for helpful comments. Support for this research was provided by the National Pharmaceutical Council. We disclose the use of Generative AI in developing the analysis code.

[†]Risk & Insurance, Wisconsin School of Business, University of Wisconsin - Madison (dan.sacks@wisc.edu).

[‡]Weatherhead School of Tropical Medicine and Public Health, Tulane University (brigham.walker@tulane.edu).

1 Introduction

Since the COVID-19 pandemic, vaccination rates have fallen across a wide category of diseases. The share of kindergartners not vaccinated against measles, mumps, and rubella increased from 4 percent in 2020 to 6 percent in 2025; polio and hepatitis-B vaccination rates fell similarly (Centers for Disease Control and Prevention, 2025). Adult vaccination rates have fallen as well, with flu vaccination rates falling each year since 2020 (Centers for Disease Control and Prevention, 2026). These declining vaccination rates are concerning given the health costs of vaccine-amenable diseases and the widespread scientific evidence supporting the safety and effectiveness of vaccination.

Many interventions have been tested to address declining vaccination rates, but with limited success, especially for vaccine hesitant people. For example, text-message based nudges and reminders sent to patients who had gotten the flu vaccine in prior seasons raised flu vaccination rates by 1-3 percentage points (Milkman et al., 2022), and COVID-19 vaccine booster reminders had a similar-sized impact on among those who had completed their initial vaccination course (Chang et al., 2023). Reminders ahead of primary care visits also increase vaccination by 1-3 percentage points (Patel et al., 2023). In Sweden, incentives for the COVID-19 vaccine raised first-dose vaccination rate by 4 percentage points on a high base rate, and booster incentives raised second dose rate by 9 percentage points (Campos-Mercade et al., 2024). These moderate effect sizes show that reminders, though helpful, are not sufficient to achieve high vaccination rates. Effects are smaller among vaccine hesitant people. For example, reminders to vaccinate—with or without financial incentives—targeted to patients who had gotten their flu vaccine by January (late in the flu season) produced no effect on vaccination.

The limited success of nudges, reminders, and incentives indicates that relatively few vaccine hesitant people are close to the margin of vaccinating. Increasing vaccination rates among vaccine hesitant people may therefore require changing their beliefs about the costs and benefits of vaccinating, and in particular addressing excessively negatively beliefs. In practice changing these beliefs has proven challenging. Addressing beliefs via information provision has yielded mixed results: dispelling myths did not increase flu or pediatric vaccination rates (Nyhan et al., 2014; Nyhan and Reifler, 2015), while highlighting harms of pediatric illness raised support for pediatric vaccinations (Horne et al., 2015). Endorsements from trusted messengers may help (Larsen et al., 2023; Alsan and Eichmeyer, 2024), albeit with modest effect sizes.

In this paper, we test whether providing scientific evidence about flu vaccine side effects to vaccine hesitant people can shift beliefs about their own risks from the flu vaccine. By scientific evidence, we mean results from clinical trials and academic studies of vaccines. Scientific information provides part of the basis for the widespread recommendation to evaluate—establishing the safety of vaccines—so it is natural to communicate. However it may not influence beliefs if vaccine hesitant participants do not trust the information source or do not believe it is relevant to them. This is a potentially important concern

because a characteristic of vaccine hesitancy is skepticism of government and doctor advice.[1] We therefore investigate whether the source of the information—industry trial or academic study—influences belief updating, and we ask whether side effect information can shift intentions to vaccinate and actual vaccinations. Existing interventions have not specifically tested providing scientific information about short run side effect risks. For example, messaging campaigns emphasize the safety and effectiveness of vaccines (e.g. Athey et al. (2023); Alsan and Eichmeyer (2024)) but do not provide quantitative information, explain the source of the information, or test specifically the role of side effect information.

We focus on the flu vaccine because flu remains an important seasonal disease, killing tens of thousands of Americans annually and infecting tens of millions. We provide side effect information because prior work has shown that flu vaccine hesitancy is strongly associated with pessimistic beliefs about side effects (Chapman and Coups, 1999; Tsutsui et al., 2012). Quantitative belief elicitations show that vaccine hesitant people believe the flu vaccine side effect rate is 7-15 percentage points on average, much higher than clinical trial reports (Sacks and Sydnor, 2025; U.S. Food and Drug Administration, 2025, 2019). Thus there is scope to move side effect beliefs with scientific information. Flu vaccine hesitancy is also associated with more negative views about flu vaccine effectiveness and side effect risk, but those views are nonetheless consistent with CDC estimates (Sacks and Sydnor, 2025).

We use an online experiment with 3,526 vaccine hesitant participants to estimate the effects of flu vaccine side effect information provision. In our control condition, we inform participants that vaccine manufacturers use placebo-controlled clinical trials to establish the effectiveness and safety of vaccines, and we tell them that the severe adverse event rate in the placebo arm of a flu vaccine trial was 3 percent. In our three treatment conditions, we provide information about the low adverse event rate in the vaccine arms. One condition describes results from the industry-sponsored trial. A second condition reports the low adverse event rate in an academic study, and a third emphasizes the representativeness of participants in the academic study to current participants. We describe this as scientific information provision because we describe the setup and results of research studies. We measure beliefs prior to the intervention and, in a modified form, after the intervention. We measure beliefs both about the trial adverse event rate and about personal side effect risk, our primary outcome. These belief measures are correlated with proxies for vaccine hesitancy (such as trust in doctors or willingness to follow government recommendations), as well as with prior experiences with the flu and COVID vaccines.

Descriptive evidence suggests a role for scientific information provision to help address vaccine hesitancy. Vaccine hesitant participants are less likely to obtain health care information from traditional sources such as their doctor, the CDC, university, or news, but more likely to obtain information from recently emerging sources such as social media and podcasts. All participants, regardless of vaccine hesi-

[1] We show this below and willingness to follow doctor or government recommendations is a component of vaccine hesitancy scales (Larson et al., 2015).

tancy, view social media and podcasts as less reliable, suggesting they may be less likely to convey objective information. Thus it is possible that vaccine hesitant participants are less aware of the evidence base establishing the safety of the flu vaccine. Vaccine hesitant people also hold pessimistic side effect beliefs: in our control group, the average vaccine hesitant participant believes that vaccinating will increase their experience of severe adverse events by 15 percentage points, but clinical trials indicate an effect of 0-1 percentage points (U.S. Food and Drug Administration, 2025, 2019). These beliefs are highly correlated with prior vaccine experience (especially self-reported experiences of severe events post vaccination), and beliefs about personal risk are correlated with beliefs about risk in the clinical trial, which we elicit in an incentivized manner.

We find that providing scientific information reduces side effect beliefs and raises vaccine intentions among the vaccine hesitant. Providing information about an industry study side effect rate reduced personal side effect beliefs by about 3 percentage points, 20 percent of the control mean. We find larger effects for people with more pessimistic prior beliefs, and for people who find university-provided information more reliable. These findings suggest that participants indeed update their belief from the provided information. We find a similar effect size of providing information about low side effects in an academic study. Thus although study source may matter, the academic/industry distinction is not the most salient source of difference. We find no effect on beliefs of an intervention highlighting the personal relevance of the provided information, but this intervention was ineffective in the sense that participants did not view the information as especially relevant or trustworthy. Our academic arm (but not industry arm) also increased intention to vaccinate, which grew by 3 percentage points (45% of control mean) in the academic arm. However, this increased intention did not translate into a statistically significant increase in vaccinations (measured two weeks later in a follow-up survey), as we lacked the power to detect plausible increases in vaccinations.

Our results therefore show that the academic- and industry-based studies produced similar belief responses but different responses for intentions. This puzzling result is largely explained by trust in the described industry trial, which turns out to be a key moderator. People with higher trust respond strongly on both beliefs and intentions. Participants with low trust respond only slightly on beliefs, and for low trust participants we predict a *negative* effect on intentions, a backlash effect similar to that documented by Nyhan et al. (2014). This backlash effect at low trust levels generates a zero overall effect of the industry-based study information on intention to vaccinate.

We note three limitations of our findings. First, because belief responses were not especially large, we lacked the power to detect effects on actual vaccinations. Second, our participants are recruited from an online panel not necessarily representative of the US as a whole. Third, our main outcome measures—beliefs and intentions to vaccinate—are subjectively assessed. This raises the possibility that expressive answering or experimenter demand explain our findings. Although we cannot completely rule out this

interpretation, we address it in multiple ways. It is difficult for expressive answering or experimenter demand to fully explain our findings. For example, to account for our pattern of heterogeneous effects, it would have to be true that high trust participants are also especially sensitive to perceived experimenter demand, and our treatment arms differentially signal our intentions. Further, beliefs about trial outcomes both respond to treatment and are highly correlated with beliefs about personal outcomes, a pattern inconsistent with expressive answering. To additionally address experiment demand, we asked participants what they thought the study was about, and our results are robust to excluding the 10 percent of participants who understood our intent.

Overall, our results point to a potentially promising class of interventions. Providing scientific information about side effects appears effective at moving beliefs and flu vaccine intentions among the vaccine hesitant. Although vaccine hesitancy is characterized by mistrust of doctors and government recommendations, it is nonetheless possible to move some participants' beliefs with scientific information. As the updating happened among people who trusted the trial information, our results suggest that side effect information provision would be especially effective if delivered via a trusted messenger, or alternatively if the information provision were targeted to participants who are likely to find the source trustworthy. Prior work such as Alsan and Eichmeyer (2024); Larsen et al. (2023) has indeed shown the importance of the messenger; our contribution is to find a promising message.

2 Experiment Design

2.1 Preregistration

The analysis plan was pre-registered on AsPredicted; see `https://aspredicted.org/5p4dq2.pdf`. The preregistration describes our main hypotheses, all interventions, primary and secondary outcomes, and statistical methods. When we conduct additional analyses beyond those described in the preregistration, we describe them as such.

2.2 Trial protocol

We collected data over three waves, indicated in Figure 1. All participants were recruited via the online platform Prolific and participants took surveys on Qualtrics. The study was advertised as an academic survey asking about experiences, views, and backgrounds. We avoided signaling our intentions or study goals. Our baseline wave measured flu vaccine intentions, and we collected information on health conditions predicting vulnerability to flu, as well as several covariates: prior vaccine experiences, trust in government vaccine information, whether participants follow doctor's advice on vaccines, and the use and reliability of several sources for obtaining health care information (such as their doctors, the CDC, newspapers,

and social media).

Our main survey wave, run the week after the conclusion of our baseline survey, was administered to baseline wave participants who indicated flu vaccine hesitancy, meaning they said they "may or may not" or "do not intend" to get the flu vaccine. We think of these two responses as indicating either moderate or strong vaccine hesitancy. In the main survey, we randomly assigned participants to treatment arms, elicited prior and posterior beliefs about side effect risk, asked about intentions to vaccinate, and asked whether participants found the described study trustworthy and (separately) relevant to their vaccination decision. We gave participants the opportunity to click on a link to sign up for a vaccine appointment. After our main survey was running for two days, we noticed that enrollment was lower than anticipated in our sample size calculations, so we increased the payment. In our follow up survey (open to all main survey participants), begun a week after the main survey ended, we asked if participants had vaccinated against the flu in the last month. All survey waves included an attention check.

2.3 Interventions

We randomly assigned participants to a control condition or one of three information provision conditions. Randomization was conducted using the randomizer tool on Qualtrics. All participants are first told the following:

> Getting the flu vaccine is a personal decision. It carries benefits as well as risk. The main benefit is reducing the chance of getting severe flu.
>
> The main risk is an adverse event such as headache, muscle ache, tiredness, joint pain, fever, chills, or general discomfort, in the days after vaccination. "Severe" adverse events are ones that are bad enough to disrupt daily activity.

We focus on severe events because they represent a substantial cost - a loss of work or leisure - and are measured in clinical trials.

Participants are then told that, to measure benefits and risks, vaccine manufacturers conduct clinical trials with a placebo arm and a vaccine arm, and in the clinical trial for Fluarix, the placebo arm had a severe adverse event rate of 3 percent in the next few days (U.S. Food and Drug Administration, 2019). Specifically, this is the grade 3 adverse reaction rate fir the placebo arm, cumulating adverse reactions across all reported reaction types. The exact definition of a grade 3 reaction varies by reaction type; for pain, muscle aches, fatigue, arthralgia, and shivering it is defined as preventing normal activity, corresponding to our general definition. For redness and swelling the definition is $>$50 mm and for fever it is $> 102.2°F$. The control arm sees no further information. In the treatment arms, participants see additional information on the same screen. We test three treatment arms:

1. **Industry arm:** Participants are truthfully told that people getting the flu vaccine in the trial were not statistically more likely to experience a severe adverse event than people getting the placebo.

2. **Academic arm:** Participants are truthfully told about a study conducted by "academic researchers" in which participants were offered small rewards as an encouragement to vaccinate. They are further told that the study found no significant increase in severe adverse events, and the researchers concluded that the flu vaccine did not increase the rate of severe adverse events. (The study is Sacks and Sydnor (2025).)

3. **Representative arm:** The representative arm was identical to the academic arm except that participants were additionally told, truthfully, that the study was conducted among Prolific participants who weren't sure they wanted to get the flu vaccine.

Appendix Figure A.1 provides screen shots showing what participants see in each arm.

Rationale: Our intervention aimed to provide quantitative information on flu vaccine side effects from scientific studies. We expected that participants had a general understanding that pharmaceutical companies and researchers conduct trials to measure efficacy and side effects. By providing this background information to the control arm, we keep their experience similar to the treatment arms, and indeed the only difference between control and industry arm is one additional statistic. Our industry arm reports on the Fluarix trial, because its vaccine arm experienced especially low adverse event rate.

In the academic arm, we presented information from Sacks and Sydnor (2025) because participants in that study were representative of flu vaccine hesitant prolific participants, and hence we could truthfully describe that study in the personal arm. Our academic and personal arms could have proceeded without first describing the placebo rate in the Fluarix trial. However presenting this information let us measure information salience in a consistent way across arms, as we explain below.

Our study design decisions followed the recommendations in Haaland et al. (2023) and Stantcheva (2023). For example, we elicit priors as well as posteriors to facilitate heterogeneous effect estimation (as expected in information provision experiments). We recruit subjects using neutral language to avoid telegraphing our intentions, and throughout the survey we provide face-saving language, especially emphasizing that vaccination carries costs and benefits, to reduce experimenter demand effects. Our specific interventions were influenced by information provision experiments, such as Chopra et al. (2025) and especially Roth et al. (2024), where participants are provided with straightforward statistical information.

2.4 Measures

Primary outcome: Our pre-specified primary outcome is the short-run personal side effect belief, which we conceptualize as the believed effect of vaccination on the short run experience of side effects severe

enough to disrupt daily activities. To operationalize this measure, we first ask participants how likely they are to experience severe adverse events over the next few days, assuming they do not vaccinate and are not vaccinated.[2] Call this $\tilde{b}_i(0)$, the believed side effect rate absent vaccination. Then we ask the same question but assuming they do vaccinate tomorrow. Call this $\tilde{b}_i(1)$, the believed side effect rate given vaccination. Then the short run personal side effect belief is $\tilde{\Delta}_i = \tilde{b}_i(1) - \tilde{b}_i(0)$. This measure is person-specific, limited to a short time horizon, and a causal contrast.[3]

We measure personal beliefs twice, before and after information provision. To avoid asking the identical questions twice, our before-information-provision version uses a likert scale for $\tilde{b}_i(1)$ and $\tilde{b}_i(0)$. These prior beliefs are a covariate rather than an outcome.

Secondary outcomes: We also measure beliefs about what happened in the trial. Before providing any information, we ask all arms (including control) what they think is the severe adverse event rate in the placebo arm of the manufacturer trial. We incentivize the elicitation of this belief, offering participants an additional \$0.50 if their answer is within 1 percentage point of the truth. Our elicitation of personal and objective beliefs is similar to that in Roth et al. (2024). After the information provision screen, we ask all arms (including control) what they think is the severe adverse event rate in the *vaccine* arm of the manufacturer trial, again offering a reward of \$0.50 if their answer is within 1 percentage point of the truth. We define the side effect rate in trial as the vaccine arm rate minus the placebo arm rate. We use this secondary outcome to assess the salience of the information provided.

We collected three additional secondary outcome metrics related to vaccination behavior. At the end of the main survey we ask participants if they intend to vaccinate against the flu. We code participants as intending to vaccinate if they answer that they are already vaccinated or intend to vaccinate (other possible answers: may or may not vaccinate, intend to vaccinate). We provide links for participants to click on to schedule a vaccination, and we track whether they click on the link. Third, in our follow up survey we ask participants if they have vaccinated in the last month.

Trustworthiness and relevance: After our intervention and after we elicited personal side effect beliefs, we asked all participants how trustworthy they found the described industry study, and how relevant to their decision to vaccinate they found it. For participants in the academic and representative arm, we also asked the analogous questions about the academic study. These questions, which were drawn from Alsan et al. (2024), were on a 0-10 scale.

[2] The exact text of our question is, "For this question, please imagine, regardless of your actual vaccination status, that you **DO NOT** get the flu vaccine tomorrow or in the following few days. Thinking about yourself, how likely do you think you would be to experience a severe adverse event, such as headache, muscle ache, tiredness, joint pain, fever, chills, or general discomfort, in the following few days? (*As a reminder, severe means bad enough to disrupt your daily activity.*)"

[3] Designing the measure as a causal contrast is consistent with the recommendations of Smith et al. (2021).

2.5 Methods

Sample size determination: Our goal was to enroll 4,000 participants in the main survey, all of whom would be invited to complete the follow-up survey, yielding 700 participants per arm in the follow-up trial under an assumed 30 percent attrition rate, for 80 percent power to detect a treatment effect of 0.15 standard deviations. This is the minimum power Haaland et al. (2023) recommend for information provision experiments. Simulations indicated that we would be highly powered to detect plausible effect sizes on beliefs (e.g. 25 percent learning rates), but we would likely be underpowered to detect effects on vaccinations unless beliefs responded very sharply to our information provision (e.g. learning rates above 50 percent). We therefore pre-specified beliefs rather than vaccinations as the primary outcome. To achieve our target sample size we ran the baseline survey until we had 8,000 completions. We prespecified that we would run the main survey for 1 week or until we had 4,000 completions. We ran the follow-up survey until a month had passed from the main survey ending, because our lookback period to measure vaccination was one month.

Sample inclusion criteria: Americans 18 and older were eligible for the baseline study. All vaccine hesitant baseline survey participants were invited to participate in the main survey and all main survey participants were invited to the follow-up survey. We exclude cases who fail attention checks or with missing belief measures. For secondary outcome analysis we additionally exclude participants missing that outcome, so the sample size varies across secondary outcomes. We also excluded survey attempts after the first (a possibility we did not anticipate and therefore did not prespecify).

Specification: Our pre-specified primary specification estimates the effect of each arm on outcomes with the following linear regression, estimated by ordinary least squares:

$$y_i = \beta_0 + \beta_1 industry_i + \beta_2 academic_i + \beta_3 representative_i + X_i\theta + \epsilon_i. \quad (1)$$

Here β_1, β_2 and β_3 give the effect of industry, academic, and representative arm relative to control, and X_i is a vector of prespecified covariates, included to improve power. Specifically, we include indicators for each level of the following categorical variables: prior beliefs; baseline vaccine intentions; reported vaccine experiences, separately for flu and COVID (possible experiences: unvaccinated, vaccinated without remembered side effects, vaccinated with mild side effects, or vaccinated with moderate side effects, or vaccinated with severe side effects); demographics (age, gender, education, income, race, ethnicity), political views (from extremely conservative to extremely liberal, seven-point scale); trust in government on vaccines (ranging from strongly disagree to strongly agree), and reported health conditions (asthma, lung disease, heart disease, diabetes, kidney disease, or have a condition but prefer not to say which). We include separate indicators for missing covariate values.

Participants are humans: Large language models are capable of answering surveys in human-like

ways and avoiding some bot-detection tools, raising concerns that some participants in online panels may be AI agents rather than humans (Westwood, 2025). However, Gordon et al. (2026) conduct a study of participants in many online panels and find that less than half a percent of Prolific participants are flagged as AI Agents (using a detection tool with 100 percent sensitivity and specificity in test cases). Thus the technical feasibility of AI Agents infiltrating online surveys has not yet translated into widespread AI Agents in surveys. Prolific employs ongoing to participant quality control to mitigate this concern.[4] The current study did not employ AI detection techniques but ex post analysis (not preregistered) provides further support for the view that AI use was low. This study included multiple incentivized questions about side effect prevalence. Current LLMs consistently give the correct answer to these questions, yet only 3.6 percent of participants answered both questions correctly.

3 Results

3.1 Recruitment and summary statistics

Figure 1 reports the dates of all surveys, as well as the number of participants taking them, the number of participants eligible for and taking follow-up surveys, and the number with valid data. 8,003 people completed the baseline survey, of which 4,525 were flu vaccine hesitant and invited to the main survey, which 3,651 began, resulting in 3,526 participants after exclusions. (Sample sizes by trial arm, and counts of non-missing secondary outcomes, are in Appendix Table A.1.) All participants who completed the main survey were invited to the follow-up survey; 3,210 began the survey and after exclusions, we end-up with 2,994 completions. Attrition between the main and follow-up survey mean that we are missing vaccination data for about 15 percent of participants.

Table 1 reports summary statistics for the analysis sample, by treatment arm. The sample is well-balanced in the sense that the differences across arms are small and typically insignificant for individual covariates (and jointly insignificant). The table also indicates several important features of the sample. First the typical participant–selected for flu vaccine hesitancy–believes that a side effect is at least likely if they vaccinate, and two thirds do not intend to vaccinate at baseline (the remainder say they may or may not vaccinate). Despite their hesitancy, a majority have previously vaccinated against flu or COVID, and a substantial number indicate vaccinating and having a severe reaction. (This severe reaction rate is much higher than rates reported in clinical trials; it is possible that participants misremember their experience or misclassify their reaction.) Somewhat more than half follow their doctor's advice on vaccines, and fewer than half trust the government's recommendations on vaccinations. About one in six reports having

[4] We caution that the study, Gordon et al. (2026) was conducted by Prolific researchers, although it was also preregistered and its data are publicly available.

a health condition which would make them vulnerable to flu. Like other samples drawn from online panels, our sample is younger, better educated, and higher income than the typical American (Stantcheva, 2023).

3.2 Information sources among the vaccine hesitant and confident

Our goal is to provide scientific information—side-effect experiences in clinical trials—to vaccine hesitant Americans. This information is publicly available and disseminated through FDA-regulated inserts in vaccine packages. It is thus in principle knowable to our study participants. However, our participants may not be aware of this information if they do not get vaccine information from sources likely to repeat this information. To understand information sources, in our baseline survey we asked all participants (vaccine hesitant or not) whether they get their vaccine information from various sources. The top panel of Figure 2 shows the distribution of sources used, by vaccine hesitancy. The bottom panel shows the fraction of participants who view the source as reliable, conditional on using that source. [5]

The figure shows three important patterns. First, vaccine hesitant participants are less likely to get information from any traditional source, compared to vaccine confident participants (i.e. people already vaccinated or saying they intend to vaccinate). We call doctors, the CDC, universities, and the news traditional sources because they have existed for decades. Second, vaccine hesitant participants are more likely to report getting information from podcasts and social media, recently emerging information sources. Third, conditional on using a source, vaccine hesitant participants find the source less reliable, except for social media and podcasts.

Vaccine hesitant people engage less with traditional information sources and find them less reliable when they do engage. Thus it is plausible that they are unaware of scientific information about vaccine safety. Of course these findings. Similarly, Appendix Figure A.2 shows that vaccine hesitant participants are less likely to find the government's information about the flu vaccine reliable, and much less likely to do what their doctor recommends about vaccines.[6]

3.3 Predictors of personal side effect beliefs

Side effect beliefs are highly pessimistic compared to clinical trial outcomes. In our control group, the average participant believes 20 percent of vaccine-arm trial participants experienced a severe adverse event; in fact roughly 1.7 percent did. (lower than the 3 percent in the placebo arm). Likewise the average control group participant believes vaccinating would increase her personal side effect risk by 15 percentage points.

[5] Specifically the top panel reports the fraction of people indicating they use that source "sometimes" or "often" for health care information, and the bottom panel reports the fraction of people indicating they find the source "reliable" as opposed to "somewhat reliable" or "not reliable."

[6] These items are adapted from the vaccine hesitancy scale developed by Larson et al. (2015).

These high rates are not a consequence of a general tendency to overestimate low probabilities; they are determined by the *difference* in probabilities between vaccinated and unvaccinated states. Further, Sacks and Sydnor (2025) show that vaccine confident people estimate mean personal side effect risk to be near zero.

Despite this pessimistic bias, in our sample of vaccine hesitant participants, short run personal side effect beliefs are sensibly correlated with other beliefs and experiences. To show this, we regress personal side effect beliefs on other variables, individually and jointly; we report the results in Table 2.

The first column shows a strong but imperfect correlation between beliefs about personal side effects and beliefs about trial outcomes. Each 1 percentage point increase in belief about side effects in the trial is associated with an additional 0.6 percentage point increase in believed personal risk. That is, people who think side effects are likely in general also think they are likely to experience side effects themselves. But people also recognize that their own experience may differ from the trial one.

The next columns show the association between personal side effect beliefs and vaccine experience as well as trust in government and following doctors advice. Relative to people who never got a flu vaccine or COVID vaccine, people who experienced no vaccine reaction have lower personal beliefs. People experiencing a severe flu vaccine reaction predict they will be more likely to have severe reactions to the next vaccine. There is little extrapolation from COVID vaccine to flu vaccine conditional on flu vaccine experience. Mistrust of government vaccine advice and unwillingness to follow a doctor's advice on vaccines are strongly associated with more pessimistic beliefs, as we would expect for two variables which are signals of vaccine hesitancy.[7] The final column, which includes all predictors, shows that personal experiences load onto personal risk rather than general risk because the coefficients are similar when we condition on beliefs about trial outcomes, especially severe prior reactions.

3.4 Main experiment results show scientific information can reduce perceived risks

We present our main, pre-specified treatment effect estimates in Table 3. Column (1) shows that all treatment arms provided salient information. In our control group, the average belief about the vaccine arm severe adverse event rate is 20 percentage points. The industry arm reduced this belief by 8 percentage points, and the academic and representative arms reduced it by 4-5 percentage points.

The next column shows effects for our main outcome, personal side effect beliefs. The industry and academic arms reduce this belief by about 3 percentage points, a reduction of 20 percent of the control mean of 15 percentage points. We do not find a significant change in beliefs in the representative arm. The representative intervention appeared unsuccessful in the sense that our intention was to raise perceived

[7] Versions of these items are included on commonly used scales to assess vaccine hesitancy (Larson et al., 2015).

personal relevance, but in fact participants found this arm less relevant (Appendix Table A.2).

These estimates suggest relatively low learning rates. The learning rate is the effect of a 1 percentage point information shock on belief updating (Haaland et al., 2023). In our context, the information shock is that the side effect rate is roughly 0, the control belief is 15 percentage points, and the belief updating in the industry and academic arm is about 3 percentage points, so the learning rate is 20 percent. This is on the lower end of the range of prior estimated learning rates (Haaland et al., 2023). One reason for the low learning rate is that people do not fully extrapolate from the trial experience to their personal risk; personal risk updating as a share of the trial outcome updating varies from roughly 0 (representative arm) to 42% (industry arm) to 69% (academic arm). Another reason may be that our information was quantitatively vague. Overall, therefore, scientific information provision can shift beliefs about personal side effect risks among the vaccine hesitant, albeit incompletely.

The third column shows how the change in beliefs translated into intentions to vaccinate. The academic treatment raises vaccination intentions by 3 percentage points, suggesting that reducing side effect concerns can raise vaccine intentions. This is a statistically significant and large effect, corresponding to about 45 percent of the control mean of 6.9 percentage points. However, the industry treatment arm does not increase vaccine intentions despite reducing side effect perceptions, and the personal arm increased intentions despite not increasing the personal side effect belief. We show below that this surprising pattern of effects on intentions is explained in part by heterogeneous responses by trust in the provided information.

The final two columns show effects on vaccine behavior. Column (4) reports effects on clicking a link to schedule a vaccination. The point estimates are small and insignificant, and overall few people clicked on the link (1.5 percent in the control group). The last column shows effects on self-reported vaccination. The point estimates range from -1.4 percentage points for the industry arm to about 1.4 percentage points for the academic and personal arms. The estimates are imprecise; the 95% confidence intervals include both zero and the effect size for intentions. Thus our results are uninformative for vaccine behaviors. The imprecision here reflects limited power. Given our covariates, we are slightly under powered to detect the 3.1 percentage point effect we estimated in the academic arm.[8] Detecting the 2.4 percentage point effect size of the personal arm would have required nearly twice the sample size of the follow-up wave. We view that effect size or smaller as more plausible given incomplete follow through on intentions (e.g. Bronchetti et al. (2015)). Thus the combination of incomplete follow-through and our limited sample size can explain why we do not detect effects on vaccinations.

[8]Specifically, the residual standard deviation of vaccination in the control group in the follow-up period is 0.21, implying a sample size of 812 per arm needed for 80 percent power and an effect size of 0.031.

3.5 Heterogeneous effects by prior belief and views of university research point to belief updating

Models of belief updating predict larger effects for participants with more biased priors and for participants who find the information source more credible. Heterogeneity analyses confirm both these predictions in our data. We caution that we did not prespecify specific heterogeneity analyses.

To show larger updating for participants with worse priors, we estimate separate regressions with subsamples defined by prior side effect beliefs. Specifically we define "high" prior beliefs as believing that the severe adverse events are at least "somewhat likely" with vaccination, the median belief. We report the results in Table 4. People without high priors believe side effects are lower, and are more likely to intend to vaccinate. We estimate small and insignificant belief updating for this group. For people with high priors, however, we estimate much larger belief updating. The evidence on heterogeneous effects for intentions is ambiguous: the point estimates are larger in the high prior group for the industry and academic arm but not the personal arm, and the differences across subgroup are not statistically significant. We also find that people with diffuse priors update more; Appendix Table A.3 shows that people with no flu vaccine experience update more than people who vaccinated and experienced no reaction, although they update similarly to people experiencing mild reactions.

Table 5 reports heterogeneous effects by perceived reliability of university research. People who view university research as not reliable or somewhat reliable show little belief updating and little response of their intentions. People who view university research as reliable show much more belief updating and stronger responses in their intentions, although the intention response remains small in the industry arm and the belief response remains small in the representative arm.

These pattern are consistent with belief updating as the key driver of our results because we would expect larger belief responses when priors are further from the provided information. Of course, there is more scope for beliefs to come down when they start at a higher level. However, finding university research reliable is associated with lower baseline personal side effect beliefs, showing that the level of baseline belief is not the only factor driving belief updating.

3.6 Perceived study quality drives information effects

We now show that perceived study quality—trustworthiness and relevance of the industry clinical trial—is a key moderator of responses to the provided information. Specifically, people who view the industry trial information as high quality show substantial belief updating and increased intention to vaccinate, while people who view it as low quality show little belief updating and a reduction in their intention to vaccinate. Like our other heterogeneity analyses, this analysis was not pre-specified and so should be interpreted with appropriate caution.

We begin by noting that the average participant did not find the described studies especially trustworthy or relevant for their decision. In the control arm the average trustworthiness of the trial was 5.7 and its relevance was 5.11 on a 0-10 scale, meaning the average response was midway between "not at all trustworthy" and "totally trustworthy." We also found that information provision *reduced* trustworthiness by a statistically significant 0.3-0.4 points. (Appendix Table A.2.) Negative effects on trustworthiness are consistent with our study sample's high baseline prior; seeing information far from a prior makes that information source appear less trustworthy. However, some participants found the trial trustworthy and relevant, and they drive the belief updating and intention responses.

Next we show that trustworthiness and relevance moderate the effect of information provision. To show this, we would like to show that there is a larger treatment effect of information provision for people who viewed the study as more trustworthy or moderate. Doing so, however, requires addressing two challenges. First, our information provision reduced perceived trustworthiness and relevance (Appendix Table A.2), so we do not want to condition on these variables directly. Second, we would like to create a single dimension of trial quality capturing both trustworthiness and relevance.[9]

To solve these problems, we first combine our trust and relevance measures into a single index of perceived quality. To do so we take the first principal component of both measures, say $\widetilde{quality}_i$ for an index of perceived quality. Second, to purge $\widetilde{quality}_i$ of endogeneity, we work with a prediction given pre-randomization variables only, $\widehat{quality}_i$, obtained as the lasso using all pre-specified control variables as well as indicators for use of several information sources and perceived reliability of those sources. We select the lasso penalty parameter using 10-fold cross-validation.

We estimate heterogeneous effects by perceived trial quality with the following model:

$$y_i = \beta_0 + \left(\sum_{\text{arm } a} \beta_a 1(arm_i = a) + \gamma_a 1(arm_i = a) \cdot \widehat{quality}_i \right) + \eta \widehat{quality}_i + X_i\theta + \epsilon_i, \qquad (2)$$

where X_i are our pre-specified controls. Here the effect of arm a for a person with quality perception $\widehat{quality}$ is $\beta_a + \gamma_a \widehat{quality}$. By construction, the average effect is β_a. We plot the implied effects, by arm, outcome, and quality perception, in Figure 3, along with their 95 percent confidence intervals. (Appendix Table A.5 reports the estimated coefficients and standard errors.)

The figure shows the important role of perceived study quality. At high quality (a value of 1, the 96th percentile), the treatment effects on personal beliefs are 5-10 percentage points, and the effect on intentions to vaccinate are also large. At low perceived trial quality (a value of -1, the 8th percentile), the belief effects are near zero and the intention effects are negative. This backlash effect at low perceived quality is consistent with prior work such as Nyhan et al. (2014), who also find that providing scientific

[9]For readers unconcerned about the endogeneity of trustworthiness and relevance, Appendix Table A.4 reports models with interactions between treatment arm and perceived study quality.

information (that pediatric vaccines do not cause autism) reduces vaccine intentions among the most skeptical. The backlash is greatest in the industry arm, and this is why the industry arm produces a small average effect on intentions.

3.7 Experimenter demand does not drive our findings

Because our outcomes are self-reported, they could be potentially driven by experimenter demand. This concern is possible because our study design—eliciting beliefs both before and after providing study information—could signal to participants that we are interested in how information provision affects beliefs. Participants may consciously or unconsciously provide responses to confirm this hypothesis. On the other hand, their ability to do so is limited by the fact that we use *between* subject comparisons, i.e. comparing participants who saw different or more detailed study information. Also pushing against the experimenter demand interpretation is the fact that we incentivized our belief elicitation where possible and the incentivized and unincentivized measures are highly correlated.

Nonetheless, to fully rule out experimenter demand, we analyzed an open response question at the end of our main survey which asked participants what they thought the study was about. This set of analyses was not pre-registered. Appendix A describes our classification scheme and reports the distribution of responses. Critically, Appendix Table B.2 shows that our estimates are largely unchanged when we exclude participants who understood our study goal. We conclude that experimenter demand does not drive our findings.

4 Conclusion

We tested whether providing information from randomized trials about flu vaccine safety influences the beliefs and vaccination decisions of vaccine hesitant Americans. Providing such information is a potentially welfare-enhancing intervention because flu vaccine hesitant individuals hold pessimistic beliefs about vaccine side effects, and these beliefs appear important for their vaccination decisions. However, flu vaccine hesitancy is associated with mistrust of government's and doctor's advice, making it unclear whether information can shift beliefs or behaviors.

Our main result is that providing side effect information shift beliefs and intentions, but belief updating is limited and moderated by relatively low trust in trials. Providing information from academic and industry sponsored trials yielded similar belief updating. Emphasizing representativeness of study populations for vaccine hesitant individuals did not lead to updating. We found mixed evidence that the side effect information increased vaccinations. Information from the academic sponsored trial increased intentions to vaccinate, but information from the industry sponsored trial did not. We lacked the power to detect effects on actual vaccination.

While our results suggest side effect information can be an important part of messaging to increase vaccination, such messaging is likely best paired with a trustworthy messenger. Overall trustworthiness of the described studies was not especially high, but effects were large for participants who found the study trustworthy. Prior work such as Larsen et al. (2023) and Alsan and Eichmeyer (2024) has found that a credible messenger increases the persuasiveness of a pro-vaccine message, so an especially promising direction for future work is to pair an appropriate messenger with a message of vaccine safety.

References

Alsan, M., M. Durvasula, H. Gupta, J. Schwartzstein, and H. Williams (2024). Representation and extrapolation: evidence from clinical trials. *The quarterly journal of economics 139*(1), 575–635.

Alsan, M. and S. Eichmeyer (2024). Experimental evidence on the effectiveness of nonexperts for improving vaccine demand. *American Economic Journal: Economic Policy 16*(1), 394–414.

Anthropic (2025, October). Claude haiku 4.5.

Athey, S., K. Grabarz, M. Luca, and N. Wernerfelt (2023). Digital public health interventions at scale: The impact of social media advertising on beliefs and outcomes related to covid vaccines. *Proceedings of the National Academy of Sciences 120*(5), e2208110120.

Bronchetti, E. T., D. B. Huffman, and E. Magenheim (2015). Attention, intentions, and follow-through in preventive health behavior: Field experimental evidence on flu vaccination. *Journal of Economic Behavior & Organization 116*, 270–291.

Campos-Mercade, P., A. N. Meier, S. Meier, D. G. Pope, F. H. Schneider, and E. Wengström (2024). Incentives to vaccinate. Technical report, National Bureau of Economic Research.

Centers for Disease Control and Prevention (2025). Vaccination coverage and exemptions among kindergartners: Schoolvaxview interactive data.

Centers for Disease Control and Prevention (2026). Weekly cumulative estimated number of influenza vaccinations administered in pharmacies and physician medical offices, adults 18 years and older, united states. `https://data.cdc.gov/Flu-Vaccinations/Weekly-Cumulative-Estimated-Number-of-Influenza-Va/ysd3-txwj/about_data`. CDC data portal dataset.

Chang, T. Y., M. Jacobson, M. Shah, M. Kopetsky, R. Pramanik, and S. B. Shah (2023). Reminders, but not monetary incentives, increase covid-19 booster uptake. *Proceedings of the National Academy of Sciences 120*(31), e2302725120.

Chapman, G. B. and E. J. Coups (1999). Predictors of influenza vaccine acceptance among healthy adults. *Preventive medicine 29*(4), 249–262.

Chopra, F., C. Roth, and J. Wohlfart (2025). Home price expectations and spending: Evidence from a field experiment. *American Economic Review 115*(7), 2267–2305.

Gordon, A., D. Rothschild, F. M. Affonso, J. Sulik, D. J. Hauser, K. Pepin, and S. Jones (2026). Ai agent prevalence and data quality across multiple online sample providers.

Haaland, I., C. Roth, S. Stantcheva, and J. Wohlfart (2025). Understanding economic behavior using open-ended survey data. *Journal of Economic Literature 63*(4), 1244–1280.

Haaland, I., C. Roth, and J. Wohlfart (2023). Designing information provision experiments. *Journal of economic literature 61*(1), 3–40.

Horne, Z., D. Powell, J. E. Hummel, and K. J. Holyoak (2015). Countering antivaccination attitudes. *Proceedings of the National Academy of Sciences 112*(33), 10321–10324.

Larsen, B. J., T. J. Ryan, S. Greene, M. J. Hetherington, R. Maxwell, and S. Tadelis (2023). Counter-stereotypical messaging and partisan cues: Moving the needle on vaccines in a polarized united states. *Science advances 9*(29), eadg9434.

Larson, H. J., C. Jarrett, W. S. Schulz, M. Chaudhuri, Y. Zhou, E. Dube, M. Schuster, N. E. MacDonald, R. Wilson, et al. (2015). Measuring vaccine hesitancy: the development of a survey tool. *Vaccine 33*(34), 4165–4175.

Milkman, K. L., L. Gandhi, M. S. Patel, H. N. Graci, D. M. Gromet, H. Ho, J. S. Kay, T. W. Lee, J. Rothschild, J. E. Bogard, et al. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proceedings of the National Academy of Sciences 119*(6), e2115126119.

Nyhan, B. and J. Reifler (2015). Does correcting myths about the flu vaccine work? an experimental evaluation of the effects of corrective information. *Vaccine 33*(3), 459–464.

Nyhan, B., J. Reifler, S. Richey, and G. L. Freed (2014). Effective messages in vaccine promotion: a randomized trial. *Pediatrics 133*(4), e835–e842.

Patel, M. S., K. L. Milkman, L. Gandhi, H. N. Graci, D. Gromet, H. Ho, J. S. Kay, T. W. Lee, J. Rothschild, M. Akinola, et al. (2023). A randomized trial of behavioral nudges delivered through text messages to increase influenza vaccination among patients with an upcoming primary care visit. *American Journal of Health Promotion 37*(3), 324–332.

Roth, C., P. Schwardmann, and E. Tripodi (2024). Misperceived effectiveness and the demand for psychotherapy. *Journal of Public Economics 240*, 105254.

Sacks, D. W. and J. R. Sydnor (2025). Preferences, beliefs, and demand for the flu vaccine. Technical report, National Bureau of Economic Research.

Smith, J., A. Whalley, and N. Wilcox (2021). *Are Participants Good Evaluators?* WE Upjohn Institute.

Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics 15*(1), 205–234.

Tsutsui, Y., U. Benzion, and S. Shahrabani (2012). Economic and behavioral factors in an individual's decision to take the influenza vaccination in japan. *The Journal of Socio-Economics 41*(5), 594–602.

U.S. Food and Drug Administration (2019). Fluarix quadrivalent (influenza vaccine) injectable suspension, for intramuscular use: Package insert. `https://www.fda.gov/media/115744/download`. Accessed: 2026-03-05.

U.S. Food and Drug Administration (2025). Flulaval (influenza vaccine) injectable suspension, for intramuscular use: Prescribing information. `https://www.fda.gov/media/74537/download`. Package insert for FLULAVAL influenza vaccine; accessed 2026-03-05.

Westwood, S. J. (2025). The potential existential threat of large language models to online survey research. *Proceedings of the National Academy of Sciences 122*(47), e2518075122.

Table 1: Balance across treatment arms

	(1) Control	(2) Industry	(3) Academic	(4) Representative	p-value
Adverse event at least likely if vaccinate	0.553	0.538	0.533	0.529	0.756
Do not intend to vaccinate	0.655	0.666	0.674	0.663	0.872
No prior flu vaccine	0.400	0.376	0.440	0.407	0.055
No reaction remembered to flu vaccine	0.328	0.319	0.274	0.310	0.067
Mild reaction remembered to flu vaccine	0.207	0.247	0.236	0.219	0.194
Severe reaction remembered to flu vaccine	0.065	0.058	0.050	0.063	0.529
No prior covid vaccine	0.373	0.358	0.375	0.369	0.876
Severe reaction remembered to covid vaccine	0.107	0.092	0.084	0.106	0.291
Mild reaction remembered to covid vaccine	0.260	0.294	0.281	0.248	0.132
Severe reaction remembered to covid vaccine	0.107	0.092	0.084	0.106	0.291
Has health condition	0.169	0.218	0.168	0.182	0.033
Trust government	0.471	0.451	0.453	0.447	0.766
Follow doctor advice	0.568	0.550	0.548	0.565	0.764
Age 18–34	0.348	0.342	0.332	0.322	0.661
Age 35–49	0.400	0.385	0.418	0.389	0.495
Age 50–64	0.213	0.221	0.219	0.227	0.921
White	0.757	0.752	0.747	0.756	0.952
Hispanic	0.084	0.097	0.103	0.091	0.543
Income under 50k	0.354	0.398	0.385	0.359	0.174
College degree	0.533	0.507	0.493	0.519	0.386
Joint test					0.210
N	882	878	880	886	

Notes: Each cell reports the share of respondents in the indicated treatment arm with the given characteristic. The p-value column reports the p-value from an F-test of equality of the variable across the four arms (individual rows) or a joint F-test across all variables simultaneously (Joint test row). The sample consists of vaccine-hesistant Americans 18 and older, recruited from Prolific.

Table 2: Predictors of personal side-effect belief

	(1) Trial belief	(2) Vacc. experience	(3) Institutional trust	(4) Full model
Post-trial SE estimate	0.587			0.518
	(0.050)			(0.050)
Flu vacc: no reaction		-8.668		-5.058
		(1.782)		(1.661)
Flu vacc: mild reaction		0.286		1.640
		(2.589)		(2.605)
Flu vacc: severe reaction		19.940		17.074
		(4.747)		(4.339)
COVID vacc: no reaction		-7.509		-2.717
		(1.941)		(1.947)
COVID vacc: mild reaction		-1.487		1.043
		(2.357)		(2.164)
COVID vacc: severe reaction		1.081		1.342
		(3.337)		(3.326)
Trust govt: somewhat disagree			0.094	-1.790
			(3.864)	(3.255)
Trust govt: neither			-5.885	-5.406
			(3.802)	(3.143)
Trust govt: somewhat agree			-4.920	-4.508
			(3.807)	(3.202)
Trust govt: strongly agree			-8.183	-3.472
			(4.534)	(3.921)
Follow doctor: somewhat disagree			-5.190	-3.808
			(4.448)	(3.643)
Follow doctor: neither			-8.593	-5.329
			(4.066)	(3.290)
Follow doctor: somewhat agree			-6.617	-2.637
			(4.131)	(3.451)
Follow doctor: strongly agree			-13.140	-6.173
			(4.674)	(4.148)
Vacc intent: may or may not				-1.998
				(1.923)
N	882	882	882	882

Notes: Each column is a separate OLS regression of personal side-effect belief ($\tilde{\Delta}$) on the indicated predictors, estimated among control-arm respondents only ($N = 882$). Column 1 includes only the respondent's belief about the adverse event rate in the vaccine arm of the trial. Column 2 includes flu and COVID vaccine experience indicators (base category: no prior vaccination). Column 3 includes trust in government on vaccines and willingness to follow doctor advice (base category: strongly disagree). Column 4 includes all variables from columns 1–3 plus vaccination intention. Robust standard errors in parentheses.

Table 3: Treatment effects on beliefs, intentions, and behavior

	(1) Trial belief	(2) Personal SE belief	(3) Vacc. intent	(4) Link click	(5) Vaccinated
Industry	-7.922	-3.326	-0.000	-0.002	-0.014
	(0.853)	(1.021)	(0.012)	(0.006)	(0.011)
Academic	-4.243	-2.910	0.031	-0.003	0.012
	(0.877)	(1.007)	(0.013)	(0.006)	(0.012)
Personal	-4.661	0.386	0.024	-0.002	0.014
	(0.879)	(1.017)	(0.013)	(0.006)	(0.012)
Control mean	20.242	15.427	0.069	0.015	0.055
N	3,525	3,526	3,517	3,526	2,994

Notes: Each column is a separate OLS regression of the indicated outcome on treatment arm indicators for Industry, Academic, and Representative arms, with Control as the omitted category. Trial belief is the believed severe adverse event rate in the vaccine arm of the trial (on a 0-100 scale). Personal SE belief is the difference in believed severe adverse event rates if the participant vaccines versus if she does not vaccinate. Vacc. intent is an indicator for self-reported intention to vaccinate. Link click is an indicator for clicking on a link to schedule a vaccine appointment. Vaccinated is an indicator for reporting flu vaccination in the follow-up survey. Pre-specified controls are included but not shown. Robust standard errors in parentheses.

Table 4: Heterogeneous treatment effects by prior side-effect belief

	Low prior belief			High prior belief		
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	Trial belief	Personal belief	Vacc. intent	Trial belief	Personal belief	Vacc. intent
Industry	-4.156	-0.806	-0.015	-10.931	-5.321	0.015
	(0.998)	(0.964)	(0.020)	(1.323)	(1.693)	(0.014)
Academic	-1.429	-0.053	0.027	-6.730	-5.335	0.032
	(1.020)	(0.903)	(0.022)	(1.368)	(1.700)	(0.015)
Personal	-2.149	-0.069	0.025	-7.017	0.632	0.022
	(1.033)	(0.940)	(0.022)	(1.364)	(1.705)	(0.015)
Control mean	12.879	2.257	0.099	26.187	26.060	0.045
N	1,627	1,628	1,623	1,898	1,898	1,894
Prior	Low	Low	Low	High	High	High

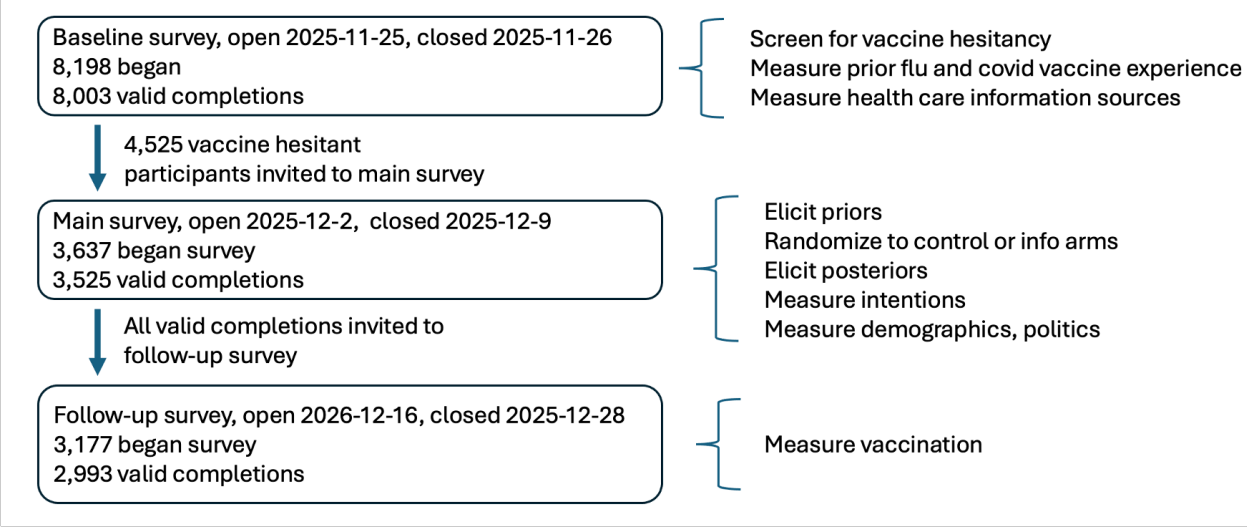
Notes: Each column is a separate OLS regression of the indicated outcome on treatment arm indicators for Industry, Academic, and Representative arms, with Control as the omitted category. Pre-specified controls are included but not shown. Robust standard errors in parentheses. The sample is split by whether the respondent's prior belief about the likelihood of experiencing a severe adverse event with the flu vaccine is below the median (columns 1–3, "Low prior belief") or at or above the median (columns 4–6, "High prior belief"). See Table 3 for outcome definitions.

Table 5: Heterogeneous treatment effects by trust in university research

	University research somewhat or not reliable			University research reliable		
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	Trial belief	Personal belief	Vacc. intent	Trial belief	Personal belief	Vacc. intent
Industry	-7.350	-0.758	-0.010	-9.320	-8.189	0.012
	(1.143)	(1.300)	(0.013)	(1.195)	(1.695)	(0.024)
Academic	-3.507	-2.345	0.012	-5.835	-4.380	0.061
	(1.168)	(1.317)	(0.014)	(1.298)	(1.565)	(0.025)
Personal	-4.161	0.896	0.004	-5.831	-1.007	0.059
	(1.180)	(1.321)	(0.013)	(1.256)	(1.617)	(0.026)
Control mean	21.762	16.137	0.053	17.506	14.149	0.099
N	2,270	2,271	2,264	1,255	1,255	1,253
Uni reliable	No	No	No	Yes	Yes	Yes

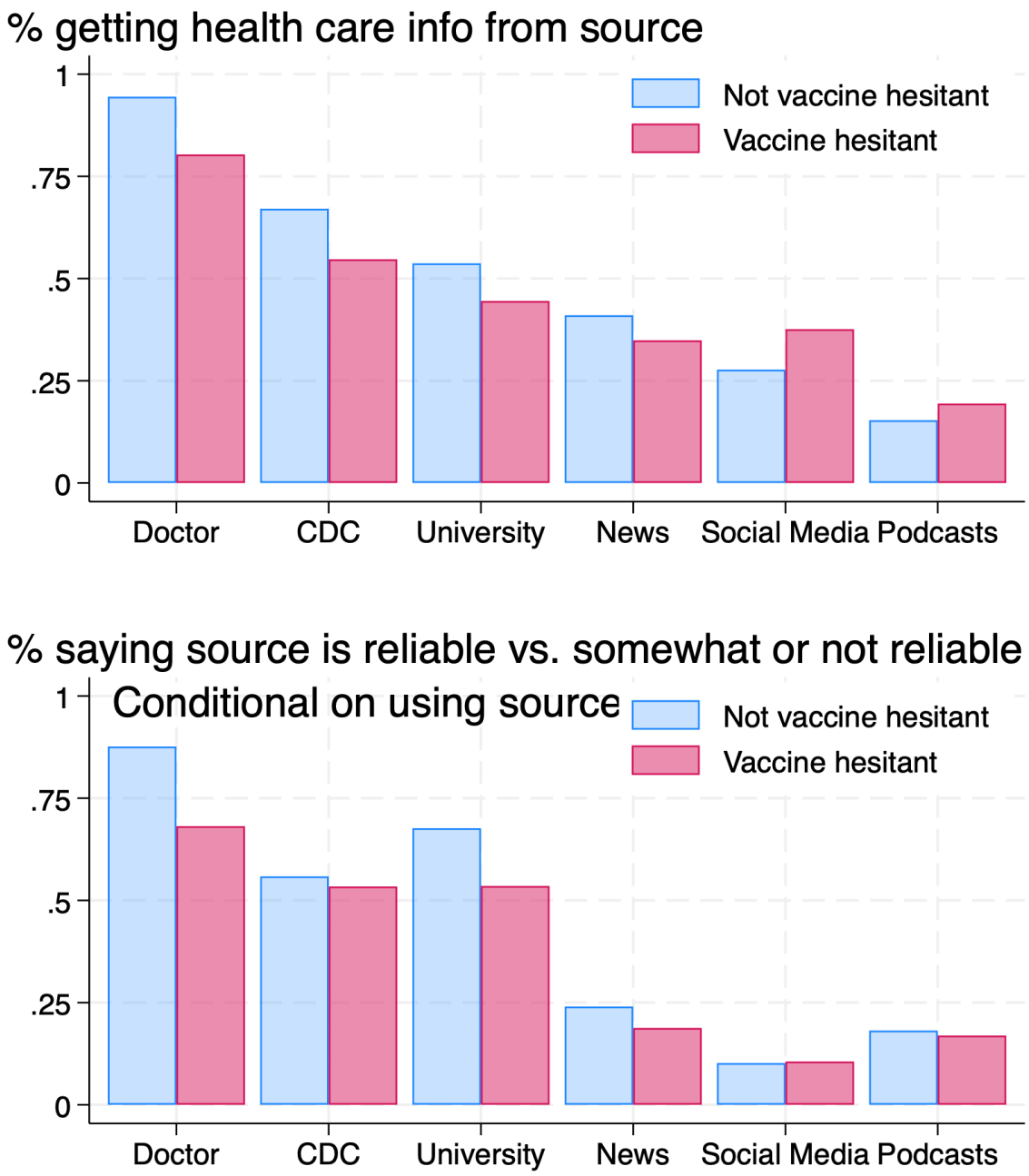
Notes: Each column is a separate OLS regression of the indicated outcome on treatment arm indicators for Industry, Academic, and Representative arms, with Control as the omitted category. Pre-specified controls are included but not shown. Robust standard errors in parentheses. The sample is split by whether the respondent reports that university research is a reliable source of health information (columns 4–6) or not (columns 1–3). See Table 3 for outcome definitions.

Figure 1: Survey waves



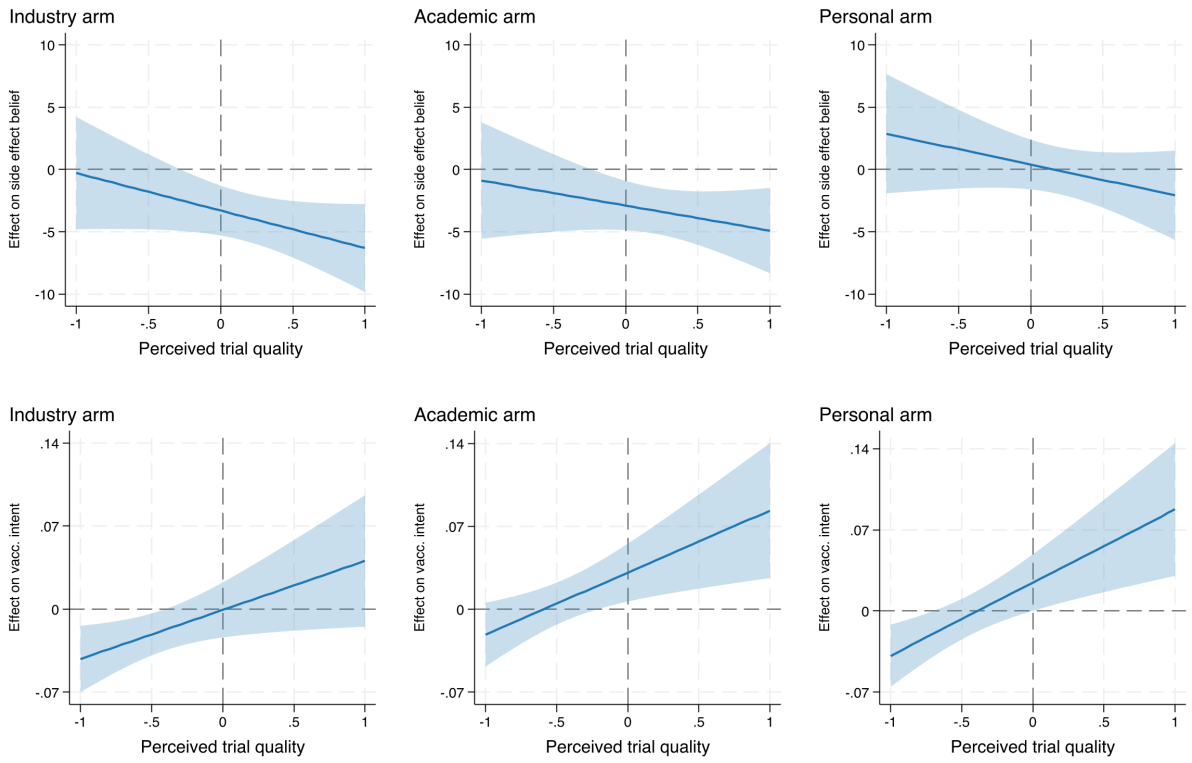
Notes: Figure reports the timing and sample sizes of each survey wave.

Figure 2: Information sources by vaccine hesitancy



Notes: Top panel reports s the fraction of vaccine hesitant and non-hesitant participants who report using the indicated source "sometimes" or "often" for health care information. Bottom panel reports the fraction who view the source as "reliable" (as opposed to "somewhat reliable" or "not reliable"), conditional on using that source. .

Figure 3: Belief updating and intent effects by perceived trial quality



Notes: Figure reports the effect of industry arm, academic arm, or representative arm on beliefs (top row) or vaccine intentions (bottom row), by perceived trial quality. Effect sizes are based on Equation 2, which assumes treatment effects are linear in quality. See text for details. Shaded area is 95% confidence interval.

Table A.1: Sample counts by treatment arm and outcome

	(1) Personal SE belief	(2) Trial belief	(3) Vacc. intent	(4) Link click	(5) Vaccinated
Control	882	882	879	882	743
Industry	878	878	876	878	755
Academic	880	880	876	880	746
Representative	886	885	886	886	750
All	3,526	3,525	3,517	3,526	2,994

Notes: Each cell reports the number of main-survey respondents in the indicated treatment arm with non-missing values for the indicated outcome. Column 3 (vaccination intention) has slightly fewer observations due to missing responses to the intention question. Column 5 (vaccinated) reports follow-up survey respondents, reflecting attrition between the main and follow-up surveys. See Table 3 for outcome definitions.

Table A.2: Treatment effects on trust and relevance perceptions

	(1) Trust in trial	(2) Relevance of trial	(3) Trust in acad. study	(4) Relevance of acad. study
Industry	-0.449	0.033		
	(0.102)	(0.134)		
Academic	-0.307	-0.114		
	(0.100)	(0.133)		
Representative	-0.417	-0.248	-0.136	-0.252
	(0.100)	(0.135)	(0.103)	(0.137)
Control/Academic mean	5.751	5.111	5.957	5.046
N	3,521	3,521	1,763	1,763

Notes: Each column is a separate OLS regression of the indicated outcome on treatment arm indicators, with pre-specified controls included but not shown. Robust standard errors in parentheses. Outcomes are measured on a 0–10 scale. Columns 1–2 use all four arms with Control as the omitted category. Columns 3–4 are restricted to Academic and Representative arm respondents, as the academic study trust and relevance questions were only asked of those arms; Academic is the omitted category so only the Representative arm coefficient is reported. The base mean for columns 3–4 is the Academic arm mean.

Table A.3: Treatment effects on personal side-effect belief by prior flu vaccine experience

	(1) No vaccine	(2) No reaction	(3) Mild reaction	(4) Severe reaction
Industry	-3.619	-1.852	-3.708	3.115
	(1.725)	(1.251)	(2.452)	(8.124)
Academic	-3.291	0.068	-5.955	-3.643
	(1.656)	(1.492)	(2.416)	(6.673)
Representative	0.687	0.726	-0.385	-5.623
	(1.741)	(1.296)	(2.519)	(6.498)
Control mean	17.406	6.789	18.101	38.386
N	1,431	1,085	802	208
Flu vacc experience	No vaccine	No reaction	Mild	Severe

Notes: Each column is a separate OLS regression of personal side-effect belief ($\tilde{\Delta}$) on treatment arm indicators (Industry, Academic, Representative), with Control as the omitted category. The sample is split by prior flu vaccine experience: never vaccinated (column 1), vaccinated with no reaction (column 2), vaccinated with a mild reaction (column 3), and vaccinated with a severe reaction (column 4). Pre-specified controls are included but not shown. Robust standard errors in parentheses. The bottom row indicates the subgroup for each column.

Table A.4: Heterogeneous treatment effects by trial quality perceptions

	Personal SE belief			Vacc. intention		
	(1) Trust	(2) Relevance	(3) Trust + rel.	(4) Trust	(5) Relevance	(6) Trust + rel.
Industry arm	1.400	-0.732	0.625	-0.041	-0.041	-0.051
	(2.918)	(2.191)	(2.817)	(0.025)	(0.019)	(0.025)
Academic arm	3.059	-1.679	0.667	-0.007	0.002	-0.010
	(3.052)	(2.142)	(2.884)	(0.026)	(0.018)	(0.024)
Representative arm	6.949	1.639	4.662	-0.027	0.004	-0.019
	(3.032)	(2.101)	(2.826)	(0.026)	(0.019)	(0.025)
Quality score	0.182	-0.482	-0.190	-0.001	0.001	0.000
	(0.356)	(0.261)	(0.179)	(0.004)	(0.003)	(0.002)
Industry arm × quality	-0.874	-0.498	-0.385	0.008	0.008	0.005
	(0.459)	(0.361)	(0.232)	(0.005)	(0.004)	(0.003)
Academic arm × quality	-1.078	-0.247	-0.346	0.007	0.006	0.004
	(0.481)	(0.358)	(0.240)	(0.005)	(0.004)	(0.003)
Representative arm × quality	-1.214	-0.277	-0.430	0.010	0.004	0.004
	(0.478)	(0.359)	(0.236)	(0.005)	(0.004)	(0.003)
Control mean	15.427	15.427	15.427	0.069	0.069	0.069
N	3,521	3,521	3,521	3,517	3,517	3,517

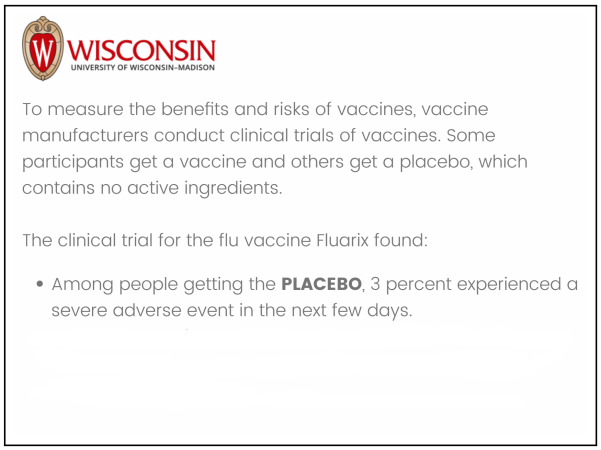
Notes: Each column is a separate OLS regression of the indicated outcome on treatment arm indicators (Industry, Academic, Representative), a measure of reported trial quality, and interactions of each arm indicator with that quality measure. Columns 1 and 4 use reported trust in the trial information (0–10 scale); columns 2 and 5 use reported relevance of the trial information (0–10 scale); columns 3 and 6 use the sum of trust and relevance (0–20 scale). These quality measures are collected after the treatment is administered. The "Quality score" row reports the coefficient on the quality measure; interaction rows report the coefficient on the arm-by-quality-score interaction. Pre-specified controls are included but not shown. Robust standard errors in parentheses. See Table 3 for outcome definitions.

Table A.5: Heterogeneous treatment effects by predicted trust and relevance

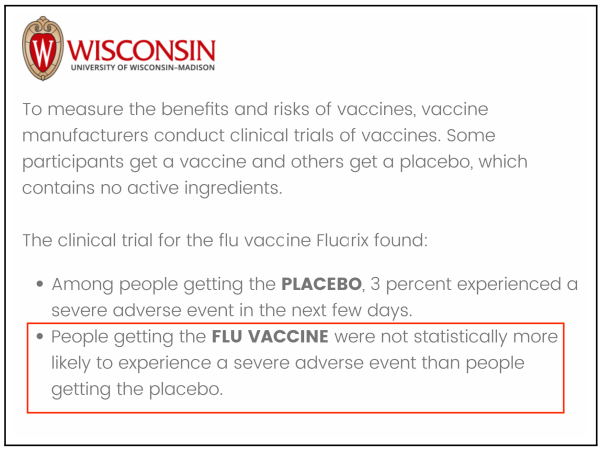
	(1) Personal SE belief	(2) Vacc. intention
Industry arm	-3.302	-0.001
	(1.024)	(0.012)
Academic arm	-2.900	0.031
	(1.010)	(0.012)
Representative arm	0.393	0.024
	(1.021)	(0.012)
Predicted quality	-1.322	0.041
	(1.978)	(0.023)
Industry arm × predicted quality	-3.009	0.041
	(1.793)	(0.019)
Academic arm × predicted quality	-2.011	0.052
	(1.832)	(0.019)
Representative arm × predicted quality	-2.473	0.063
	(1.902)	(0.019)
Control mean	15.427	0.069
N	3,526	3,517

Notes: Each column is a separate OLS regression of the indicated outcome on treatment arm indicators (Industry, Academic, Representative), a predicted trust/relevance index (`pca1_hat`), and interactions of each arm indicator with the index. `pca1_hat` is the first principal component of post-treatment trust in the clinical trial and perceived relevance of the trial, predicted from pre-experiment covariates using lasso selected in the control group. The "Predicted quality" row reports the coefficient on `pca1_hat`; interaction rows report the coefficient on the arm-by-index interaction. Pre-specified controls are included but not shown. Robust standard errors in parentheses. See Table 3 for outcome definitions.

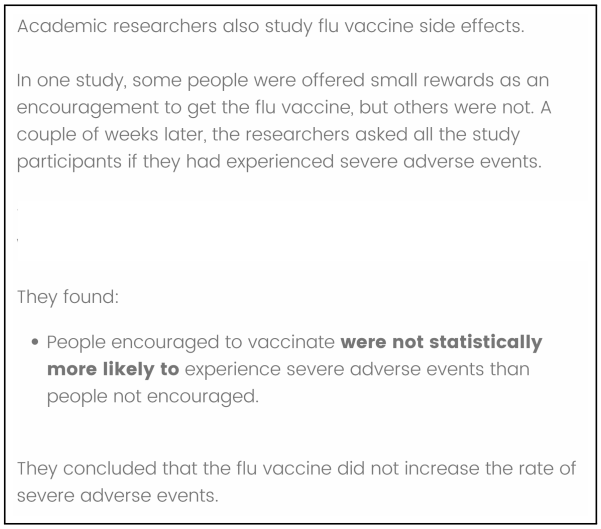
Figure A.1: Experiment conditions



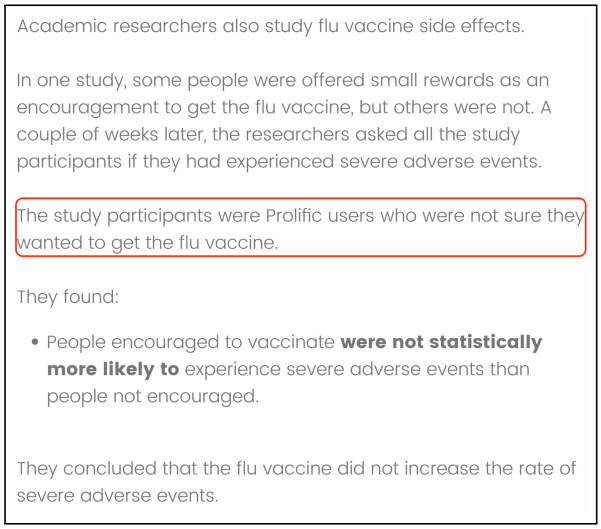
(a) Control



(b) Industry



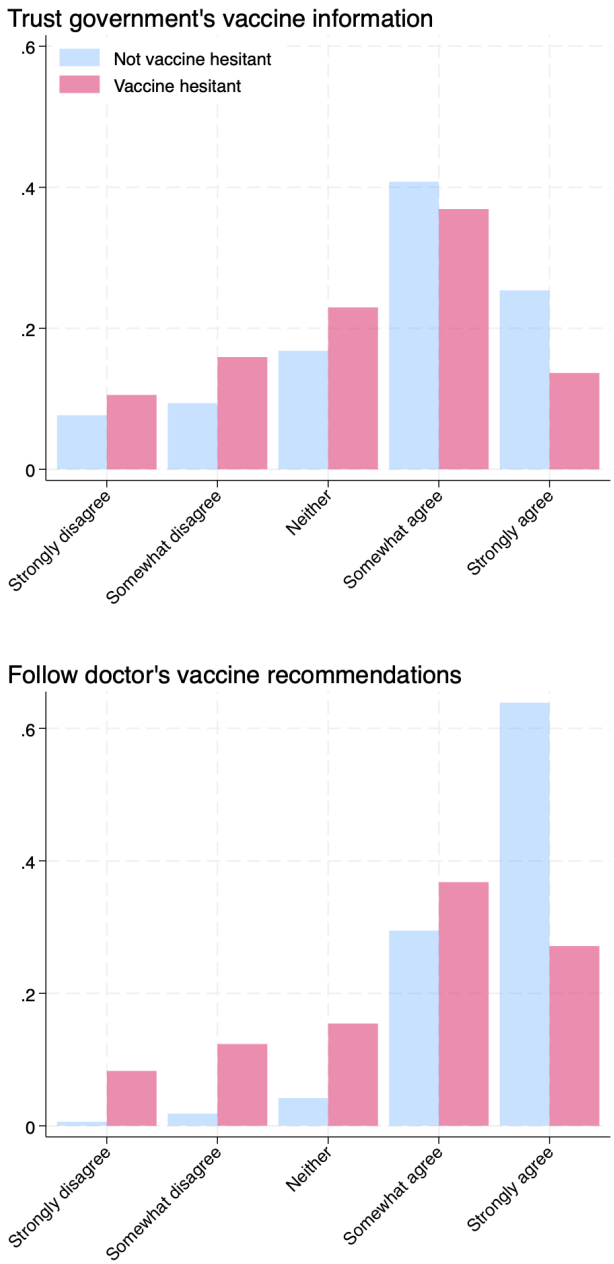
(c) Academic (additional info beyond control)



(d) Representative (additional info) beyond controls

Screenshots of the information screens for all four treatment arms. Participants in the academic and representative arms saw the control arm information as well as the information about the academic study. The red box was not included in the materials participant saw but is included here to highlight the differences.

Figure A.2: Trust in government and following doctor's advice by vaccine hesitancy



Notes: Top panel reports s the fraction of vaccine hesitant and non-hesitant participants who report using the indicated source "sometimes" or "often" for health care information. Bottom panel reports the fraction who view the source as "reliable" (as opposed to "somewhat reliable" or "not reliable"), conditional on using that source.

A Participant Perceptions of Study Intent

We investigate participant perceptions of study intent. On the last screen of the main survey, participants were asked "In a few words, what do you think was the goal of th study?" We categorized these responses to determine whether participants recognized our study goal, which was to see whether and what type of side effect information could change beliefs about side effects and intent in vaccinating.

After manually reviewing a subset of responses, we settled on the following classification scheme. Responses are assigned to one or morenot mutually exclusive categories based on mentioning the following topics: the flu or flu vaccine; attitudes, opinions, or view points; information provision or research, including clinical trials; side effects or effectiveness; intentions or plans to vaccinate; changing behavior; and unclear or unclassifiable. We said a participant understood the study if their response mentioned flu or vaccine, information, changing behavior, and at least one of attitudes, side effects or effectiveness, or intentions. Appendix Table B.1 provides examples of responses that fall into exactly one category as well as responses with multiple categories, and a response we classify as understanding study intent. Our approach to classification was guided by the ideas and examples in Haaland et al. (2025).

We used a large language model to implement this coding scheme. Specifically, we sent a prompt to Claude-Haiku-4.5 via its API (Anthropic, 2025). (The exact prompt is reported at the end of this section.) We used the lightweight Haiku model because we expected that classifying a small number of words did not require extended reasoning or other advanced capabilities. We set the temperature parameter to 0 to minimize randomness in classificaiton. We sent each participant's response separately to avoid any cross-response contaimination in the classification.

Appendix Figure B.1 reports the share of responses in each category. A large majority of participants correctly recognized that the study was about the flu vaccine, consistent with generally high attention among Prolific participants. Many participants also mentioned that it was about attitude, opinions, or viewpoints on the flu vaccine, and many also mentioned risk. However relatively few mentioned changing behavior and so overalll about 10 percent of responses indicated that participants understood the study's intent.

To show that these participants are not driving our results, we re-estimate our main model but excluding anyone classified as understanding the study's intent. The results in Appendix Table B.2 are quite similar to our main results. The estimted effects on beliefs are all within about 0.2 percentage poitns of the baseline estimate. The estimated effects of the industry and academic arm on intention are essentially identical. The effect of the personal arm on intention is about a quarter smaller. The effect on linked click remains precisely zero and the effect on vaccination is also essentially idetical (slgihtly larger effect of academic arm, sligtly smaller effect of personal arm).

Table B.1: Example Responses by Classification Pattern

Classification pattern	Example response
Mentions flu or vaccine only	"Flu vaccines"
Mentions feelings, attitudes, or opinions only	"How people react to placebos"
Mentions information or research only	"Paying attention to the study by reading the information given by the researcher."
Mentions side effects or effectiveness only	"I am not sure. Maybe taking risks."
Mentions vaccination intentions only	"To see what decisions I will make."
Mentions changing behavior, attitudes, or beliefs only	"To see if you could be influenced"
Mentions flu/vaccine and feelings/attitudes	"The goal of the study was to evaluate participants perceptions of flu-shots."
Mentions flu/vaccine and information	"Information about vaccinations"
Mentions flu/vaccine and side effects/effectiveness	"To see if people think flu vaccines cause adverse effects"
Mentions flu/vaccine, feelings/attitudes, and information	"perception of information and shots"
Mentions flu/vaccine and changing behavior, without information	"encourage flu vaccine"
Understood study's intent (flu/vaccine, information, change, and attitudes, side effects/effectiveness, or intentions)	"To see if people would change their mind on getting the vaccine upon learning about a study that shows people getting the vaccine aren't more likely to experience symptoms than those who got the placebo."

Notes: Each row shows one participant's verbatim response to "What do you think this study was about?", illustrating a pattern of categories (defined in code/classification_prompt.md) assigned by the claude-haiku-4.5 classifier.

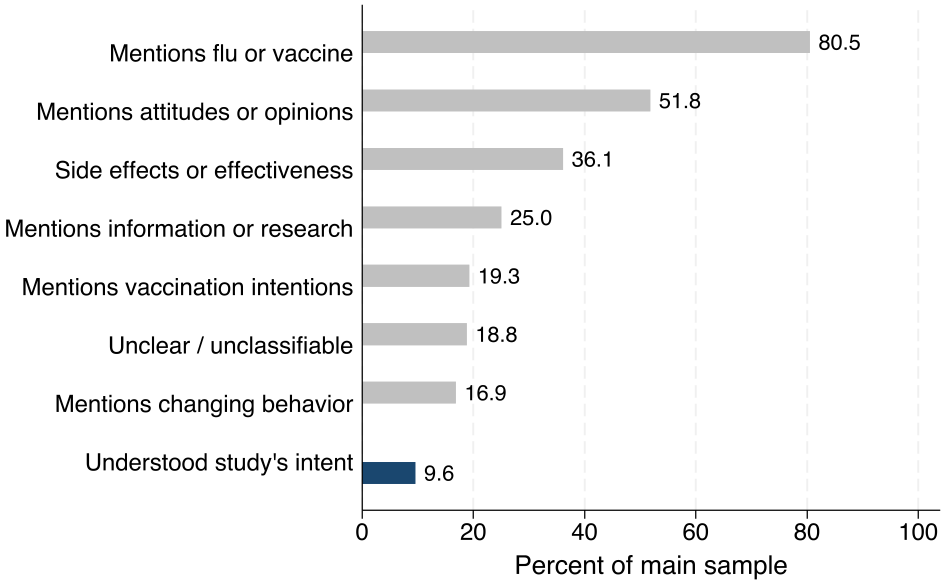


Figure B.1: Classification of Open-Ended "What Was This Study About?" Responses

Notes: Bars show, for the main sample, the percent of responses to "What do you think this study was about?" classified by claude-haiku-4.5 (prompt in code/classification_prompt.md) as mentioning each category. Categories are not mutually exclusive. "Understood study's intent" indicates the response was classified as mentioning flu/vaccine, information, and change, and at least one of attitudes, side effects or effectiveness, or intentions.

Table B.2: Treatment effects excluding respondents who understood the study's intent

	(1) SE (trial)	(2) Delta	(3) Vacc Intent	(4) Link Click	(5) Vaccinated
Industry	-7.980	-3.412	-0.001	-0.004	-0.012
	(0.901)	(1.072)	(0.013)	(0.006)	(0.012)
Academic	-3.911	-2.908	0.031	-0.002	0.015
	(0.928)	(1.072)	(0.013)	(0.006)	(0.013)
Personal	-4.563	0.530	0.018	-0.004	0.012
	(0.938)	(1.089)	(0.013)	(0.006)	(0.013)
Control mean	20.672	15.480	0.072	0.015	0.055
N	3,186	3,186	3,177	3,186	2,697

Notes: Each column is a separate OLS regression of the indicated outcome on treatment arm indicators for Industry, Academic, and Personal arms, with Control as the omitted category. Pre-specified controls are included but not shown. Robust standard errors in parentheses. The sample is restricted to main-sample respondents whose response to "What do you think this study was about?" was not classified as indicating they understood the study's intent (Figure B.1; classification described in code/classification_prompt.md). See Table 3 for outcome definitions.

Classification Prompt

The following prompt was used (via `code/classify_open_responses.py`) to classify each open-ended "What do you think this study was about?" response with claude-haiku-4.5 (`code/classification_prompt.md`):

```
You will be given a response from a survey experiment participant who
 was asked what they thought the study that they participated in was
about. The study was about the effect of information provision on flu
 vaccine beliefs and intentions. We will say a participant understood
 the study's intention if they thought the study was about vaccines,
information provision, and changing beliefs or intentions.

Your task is to interpret the response, flag whether it refers to
subjects described below, and use that to infer whether the
respondent currently understood what the study was about. Note that
the string "<NL>" in a response means a new line character.

Use semantic interpretation. If you cannot point to specific meaning
in the explanation, set unclear=1 and all other fields to 0.


For the response, provide the following classifications:
- about_flu: =1 if the response mentions flu or vaccine.
- about_attitude: =1 if response mentions feelings, views, attitudes,
 opinions, perceptions, viewpoints
- about_info: =1 if the response mentions information provision,
including research or studies. Should not =1 if the repsondent only
says the stud was about gathering information.
- about_beliefs: =1 if the response mentions risks, side effects,
effects of vaccines, or effectiveness
- about_intent: =1  if the response mentions vaccine intentions or
interest in vaccination, or whether people will vaccinate
- about_change: =1 if the response mentions changing behavior,
attitudes,  or beliefs
- unclear: =1 if response cannot be classified, including blanks,
unclassifiable, or single words. Unclear =0  if any other response is
 equal to 1.

Afterwards, we will infer understood-intent from if response includes
 flu AND change AND information AND (attitude or beliefs or intents)
```