Knowledge versus Skills: Evidence from a Field Study on Health Misinformation

Fatu Conteh Rachel Glennerster Samantha Horn Anne Karing*

Abstract

Digital communication—particularly social media—has transformed how people consume health information, making misinformation a pressing policy concern. Traditional efforts to counter misinformation rely on educational campaigns to disseminate accurate information. More recent approaches teach people to assess information accuracy from its features rather than its content. We compare these two approaches in the context of vaccine information in Freetown, Sierra Leone. Participants were assigned to a control group, a domain-specific vaccine knowledge training (*VaccineKnowledge*), or training to recognize common features of misinformation (*MisinformationSkills*). Both interventions improve vaccine information discernment: *VaccineKnowledge* improves accuracy only for covered content, while *MisinformationSkills* yields broader, generalizable improvements. *MisinformationSkills* also improves selective sharing, reducing false but not true headline sharing, while *VaccineKnowledge* has similar but less precise effects. These results show that building domain knowledge and misinformation detection skills are comparably effective strategies for improving information discernment and sharing behavior. However, despite these discernment gains, neither intervention leads to independent knowledge acquisition when participants encounter new vaccine-related information.

Keywords: information discernment, misinformation, vaccination, public health
JEL Codes: I12, I18, D83, C93

*Conteh: RamLABS (email: fconteh@ramlabs.org). Glennerster: Center for Global Development (rglennerster@cgdev.org). Horn: University of Chicago (email: samihorn@uchicago.edu). Karing: University of Chicago (email: akaring@uchicago.edu). We thank Gaia Gaudenzi, Tom Lanoizelee, and Anbar Aizenmann for excellent research assistance. The experiment and data collection were approved by University of Chicago IRB and the Office of the Sierra Leone Ethics and Scientific Review Committee. The experiment was registered at AEA RCT registry (RCT ID: AEARCTR-0013488). Pre-analysis plan available at `https://drive.google.com/file/d/1kVN-0xjtI6lvc1R3csTTeuC1WEtNxm4G/view?usp=sharing`. All errors are our own.

1 Introduction

Misperceptions about health are not new (Pliny, 77). Traditionally, people relied on local networks and experts for health information, constraining the reach of both accurate and inaccurate information. The shift to digital communication—accelerated by social media—has dramatically expanded access to both, particularly in many low-income countries, where social media is now a primary news source (Hagg et al., 2018). Policy efforts to counter inaccurate information often rely on the dissemination of accurate information through educational campaigns and programs. More recently, digital media literacy interventions have been developed to teach people how to discern inaccurate information in digital spaces based on its features (Guess et al., 2020; Ali and Qazi, 2023; Amar et al., 2025). However, it remains unclear how these two approaches compare—whether providing knowledge improves information discernment, and whether strengthening discernment, in turn, can lead to knowledge acquisition.

We explore these questions in the context of vaccine-related information in Sierra Leone. Several factors motivate this choice. First, vaccines are a frequent target of both accurate and inaccurate claims, particularly on social media (Larson, 2020). Second, understanding vaccines requires more technical knowledge than simpler health interventions, as many people lack a clear grasp of how they work (Larson Williams et al., 2019). This contrasts with simpler technologies like mosquito nets, where the mechanism of protection is more easily understood. Third, vaccine hesitancy is a growing global health challenge, making it an important policy setting for studying how people assess information.

To compare the effects of building vaccine knowledge and information discernment skills, we conducted a lab-in-the-field experiment with 1,818 participants in Freetown, Sierra Leone. Participants were randomly assigned to one of three short training interventions. The first, *VaccineKnowledge*, provided training on how vaccines work and their safety—specifically, how they train the body to fight disease, what causes side effects, and the rigorous testing required for approval. The second, *MisinformationSkills*, trained participants to recognize common features of inaccurate health information—specifically, the absence of a credible source, emotionally charged language, a mix of true and false claims, and reliance on unverified experts. This approach drew on prior research (Guess et al., 2020; Roozenbeek et al., 2022; Ali and Qazi, 2023). The control group received training on an unrelated health topic—hand-washing.

We measure effects on discernment of both real and experimental vaccine-related headlines immediately after training and again one month later to assess persistence. To explore mechanisms, we test whether discernment gains reflect increased domain knowledge, the use

of misinformation detection heuristics, or shifts in general skepticism or optimism. Experimental headlines vary systematically in relation to the *VaccineKnowledge* training content, the presence of misinformation features, accuracy, and sentiment, allowing us to distinguish between these pathways. Beyond discernment, we use lab-based exercises to measure downstream information behaviors, specifically learning from and sharing new information. For consenting participants (74%), we track real-world sharing using a web-based application (Garimella and Chauchard, 2024) that records WhatsApp activity and messages between midline and endline.[1] Finally, we evaluate effects on policy-relevant health behaviors—vaccine uptake and formal versus informal healthcare utilization.

Both treatments lead to modest gains in discernment: 0.09 SD (FDR-adjusted $q < 0.01$) for *VaccineKnowledge* and 0.18 SD ($q < 0.01$) for *MisinformationSkills*. These magnitudes are comparable to or exceed those reported in prior misinformation interventions. Gains in real headline discernment are similar in magnitude, suggesting that improvements generalize beyond the experimental headlines. We remeasure discernment one month later to assess persistence. At each survey wave, headlines are randomly drawn at the participant level from the same broader pool, ensuring comparability across survey rounds, though the endline headlines are a subset of the midline ones in that they only relate to content covered in the *VaccineKnowledge* training. Both treatments continue to show positive effects: *VaccineKnowledge* improves discernment on experimental headlines by 0.21 SD ($p < 0.01$) and *MisinformationSkills* by 0.22 SD ($p < 0.01$). Both effects are directionally larger than the midline effect on the comparable set of headlines but not statistically significantly different. Real headline discernment follows a similar pattern—higher than midline results in both arms but only statistically significantly so.

Turning to mechanisms, we examine whether improvements in headline discernment stem from increased vaccine knowledge or the application of misinformation detection skills, such as source evaluation. We also explore whether gains are offset by unintended shifts toward skepticism of true information or trust in false but positively framed information.

We find that *VaccineKnowledge* improves headline discernment for *Core* content (0.16 SD, q <0.01)—topics directly covered by the training—but find no evidence of transfer to *Adjacent* content (0.02 SD, $q = 0.19$)—topics related to but not explicitly covered in the training. In contrast, *MisinformationSkills* boosts discernment broadly, generating gains of

[1]We collected 2,834,786 messages from participants. Because using large language models (LLMs) alone for broad-scale text classification can introduce non-classical measurement error, we are currently in the process of human-coding a subset of forwarded and sent messages to identify vaccine- and health-related information and misinformation. Following the estimation approach in (Ludwig et al., 2025), we will then rely on LLM-based classification for the remaining text data and use the human-coded subset to model—and thereby adjust estimates for—the potential measurement error.

around 0.16-0.17 SD for both content types (q <0.01). There is suggestive evidence that these gains primarily arise from source evaluation, though we cannot rule out contributions from other learned heuristics, such as identifying emotional language. Unlike findings from prior misinformation interventions (Hoes et al., 2024), we find no evidence of increased skepticism toward true headlines: both treatments increase discernment of true headlines. Similarly, we find no evidence of misplaced optimism—while participants in the *VaccineKnowledge* arm perform worse on false positive than false negative headlines, their discernment on false positive headlines does not significantly differ from the control group.

Together, these findings indicate that knowledge-based interventions improve discernment for trained topics but show limited transfer even to related within-domain content. In contrast, skills-based approaches yield broader, more generalizable benefits.

Next, we investigate whether discernment gains translate into more accurate beliefs following exposure to true and false headlines—that is, whether discernment facilitates knowledge acquisition. We elicited incentivized beliefs of HPV vaccine safety (the percentage of girls expected to die from the vaccine) and efficacy (the percentage of cervical cancers the vaccine can prevent) from half of the sample without additional information and the other half after exposure to a set of related headlines. This allows us to separate treatment effects on pre-exposure beliefs from those on learning.

For both questions, participants in the *VaccineKnowledge* arm report more accurate initial beliefs than those in the *MisinformationSkills* arm, though only beliefs about vaccine efficacy differ significantly from the control group. Learning dynamics vary by question. For HPV vaccine safety, exposure to headlines causes control participants' beliefs to become less accurate. Participants in the *VaccineKnowledge* arm do not exhibit this decline, while updates in the *MisinformationSkills* arm are not statistically distinguishable from either group. As a result, post-exposure beliefs remain most accurate for *VaccineKnowledge* participants, reflecting both more accurate pre-exposure beliefs and resilience against misinformation-induced accuracy declines. In contrast, updates about HPV vaccine efficacy following headline exposure are minimal ($\sim$1 p.p.) and statistically indistinguishable across treatments, leaving *VaccineKnowledge* participants' post-exposure beliefs more accurate primarily because of their more accurate initial beliefs. These findings indicate that *VaccineKnowledge* improves the accuracy of vaccine-related beliefs and mitigates inaccurate belief updating in response to misinformation, whereas discernment gains from *MisinformationSkills* do not reliably lead to more accurate belief updating when exposed to new information.

Motivated by prior work showing that discernment gains do not always translate into changes in sharing behavior (Pennycook et al., 2021), we examine effects on sharing decisions. To reflect the reputational incentives of real-world sharing, participants had the opportunity

to share vaccine-related news about the new malaria vaccine with other community members in a WhatsApp group. The *MisinformationSkills* arm improves sharing discernment—the difference between the percentage of true and false headlines shared—by 6 p.p. ($p = 0.04$). The *VaccineKnowledge* arm exhibits a similar but not statistically significant effect of 5 p.p. ($p = 0.11$), with no significant difference between the two treatments ($p = 0.71$). Given concerns that misinformation-focused interventions could reduce the sharing of true information (Guriev et al., 2023), it is notable that neither treatment significantly affects the sharing of true headlines.

Finally, we assess treatment effects on real-world health behaviors one month after the intervention, focusing on Covid-19 vaccine uptake—measured by offering the vaccine near the survey site—and self-reported use of formal versus informal healthcare providers. We find no significant effects on either outcome. However, the confidence intervals rule out increases in vaccine uptake exceeding 8 to 10 percentage points, magnitudes similar to the improvements in headline discernment documented above.

Our work makes several contributions to the recent literature on interventions designed to improve discernment of misinformation on social media and other digital platforms (Chan et al., 2017; Nyhan et al., 2020; Guess et al., 2020; Lewandowsky and Van Der Linden, 2021; Pennycook and Rand, 2022; Ali and Qazi, 2023; Kozyreva et al., 2024; Fazio et al., 2024; Amar et al., 2025). First, we provide causal evidence on the relative efficacy of building domain expertise versus training people how to recognize misinformation in improving information discernment within a specific health domain. While prior work finds a correlation between expertise and discernment ability (Nan et al., 2022), there is little causal evidence on whether increasing expertise directly improves discernment—or how it compares to direct training in misinformation detection. We show that both approaches improve information and sharing discernment, indicating that they represent distinct but viable strategies for countering false claims in public health.

Second, we add to a small but growing literature on misinformation interventions in low-income contexts where information primarily circulates through private networks rather than public online platforms like Facebook or Twitter (Guess et al., 2020; Badrinathan, 2021; Ali and Qazi, 2023; Amar et al., 2025). In such settings, widely studied policies such as fact-checking and accuracy reminders face implementation challenges. We show that two scalable, media-based interventions can improve discernment even in environments where information flows through private and informal channels, making interventions that rely on content monitoring less feasible.

Third, we examine how these interventions impact belief updating when participants encounter new information. Although discernment is expected to facilitate learning—by

helping people distinguish reliable from unreliable claims—this relationship remains largely untested. If discernment alone is sufficient, misinformation training could serve as an alternative path to domain expertise. However, we find that even when participants encounter accurate information, improved discernment alone does not reliably translate into more accurate belief updating, suggesting discernment may be necessary but insufficient for knowledge acquisition.

Fourth, while most studies measure sharing discernment through stated intentions, few assess actual sharing behavior (Guriev et al., 2023). We contribute new evidence by examining both participants' sharing decisions in a WhatsApp-based lab setting, where they share news with others, and their real-world WhatsApp activity outside the lab. WhatsApp is the dominant platform for information exchange in Sierra Leone and many other low-income countries, yet its privacy features make observational research challenging. By capturing both controlled and real-world sharing, our approach provides direct behavioral evidence on how interventions shape information transmission.

Our work also relates to the literature on targeted educational interventions, which have been shown to influence behaviors such as safer drinking water use (Madajewicz et al., 2007; Jalan and Somanathan, 2008), sexual partner selection (Dupas, 2011), and contraceptive uptake (Bau et al., 2024). While such interventions typically focus on increasing knowledge to change behavior, less attention has been paid to how these approaches shape how individuals evaluate new information, update beliefs, and decide what information to disseminate. Our findings show that a brief, scalable educational intervention on vaccination can affect not only knowledge but also how individuals critically engage with health information, suggesting a broader potential role for education in fostering resilience against misinformation.

2 Context

2.1 Vaccine Acceptance

In Sierra Leone, vaccine acceptance is high, with low levels of hesitancy comparable to those in much of Sub-Saharan Africa. Data from the Wellcome Global Monitor 2018 show that 95% of Sierra Leoneans consider vaccines safe and effective, and 97% of parents report that their children have been vaccinated (Solís Arce et al., 2021). In contrast, comparable figures in the United States are 73%, 87%, and 95%, for vaccine safety, efficacy, and parental uptake, respectively. Our survey data similarly reflect high confidence, with 88% control group participants agreeing that vaccines are safe and and 86% agreeing that they are effective. Acceptance of Covid-19 vaccination has also been relatively high, with surveys conducted

prior to the vaccine rollout showing that 75-85% of Sierra Leoneans intended to be vaccinated. A field study by Meriggi et al. (2024) in Sierra Leone provides further evidence for this support by showing that simply reducing logistical barriers in rural areas increased Covid-19 vaccine uptake by 24 percentage points.

The Covid-19 pandemic heightened policy-makers' focus on and concerns about misinformation. Efforts by UNICEF to detect misinformation in Sierra Leone during the pandemic identified numerous instances of rumors and incorrect information about vaccine safety in communities and online. These instances show that vaccine-related misconceptions can still arise despite general vaccine acceptance. Similarly, despite high average levels of trust in the safety and efficacy of vaccines, our sample holds several incorrect beliefs about fundamental aspects of vaccination. Among the control group, only 55% believe vaccines cannot treat an existing illness. A similar proportion view side effects as evidence that a vaccine is ineffective. Just 30% correctly identify as false the claim that the Covid-19 vaccine provides indefinite protection. For more extreme misconceptions, only 39% say it is definitely false that the measles vaccine can paralyze children, and 70% mistakenly believe that vaccines work by strengthening the skin against germs. Thus, both the prevalence of misinformation and baseline level of vaccine knowledge provide an opportunity to assess different approaches to improve information discernment.

2.2 Information Environment

Survey data show that WhatsApp (65%) and social media (64%) are participants' primary daily news sources, with 51% relying on both. By comparison, traditional media such as radio (22%) and television (23%) play a smaller role in daily information consumption. Despite the widespread use of digital platforms, trust in these sources remains relatively low, with fewer than 40% of participants expressing confidence in the information they provide. In contrast, institutional and formal sources such as the World Health Organization (85%), UNICEF (81%), and health workers (73%) are trusted by a majority of participants. This context highlights a dynamic information environment where informal, widely accessible channels dominate daily news consumption, yet institutional sources remain important for credibility.

3 Experimental Design

Our experiment was designed with four primary objectives. First, we sought to compare the effects of building domain-specific expertise and teaching misinformation evaluation skills on participants' ability to discern accurate from inaccurate information. Second, we aimed

to assess how these effects persist over time. Third, we aimed to compare the broader consequences of these interventions on information-related behaviors, including learning from and sharing information. Finally, we aimed to evaluate their implications for policy-relevant health outcomes—specifically, vaccine uptake and formal versus informal healthcare utilization.

To meet these objectives, we conducted a lab-in-the-field experiment in Freetown, Sierra Leone, from May to August 2024, targeting recruitment of 1,800 participants.[2] Participants were randomly assigned to one of three training conditions: building vaccine-specific domain knowledge, learning tips on how to identify false health information, or, as a control, building domain knowledge for an unrelated health topic—hand-washing.

We measured participants' vaccine information discernment—defined as the ability to distinguish accurate from inaccurate vaccine information—immediately the after intervention (midline) and one month later (endline). At midline, we also measured participants' ability to learn from information about the safety and efficacy of a new vaccine, as well as their decisions to share vaccine-related information with other participants from their communities. At endline, we measured downstream health outcomes.

3.1 Recruitment and Eligibility

The study population comprised adults living in Freetown, Sierra Leone. To ensure coverage of a broad and diverse set of communities within the city while minimizing spillover risks, we selected twenty-four large, high-population communities and relocated our lab—and consequently, recruitment efforts—to a new community every three to four days. Within each community, recruitment enumerators were stationed in high-traffic public areas, such as markets and transit hubs, to identify potential participants.[3]

Enumerators screened individuals for eligibility before inviting qualified participants to attend a training session. Given the study's focus on information shared via social media—specifically WhatsApp—eligibility was restricted to literate smartphone users who had actively used WhatsApp within the past five days. Additionally, because we were interested in the impact of these interventions on Covid-19 vaccine take-up, only individuals who had not completed their Covid-19 vaccine schedule were considered eligible. To further mitigate

[2]Sample size was determined through power calculations to achieve 80% power at a 5% significance level to detect a 6-percentage-point increase in vaccination uptake, assuming a control group mean of 10% and accounting for 15% attrition. The resulting sample size also provides sufficient power to detect a standardized treatment effect of just under 0.2 for other outcomes when measured at the participant level.

[3]Recruitment enumerators were not involved in the training or survey implementation, and study enumerators were not permitted to select specific participants to train. This separation minimized the risk that enumerators would recruit individuals they deemed easier to train or survey.

the risk of spillover effects, eligibility was also restricted to those who did not know anyone attending, or who had attended, training sessions conducted by our research team. For more detail on the eligibility funnel, see Appendix Table A1.

3.2 Treatments

Each training intervention followed the same format: participants first watched a short instructional video on a tablet two times, answered reinforcement questions delivered by an enumerator to aid retention and comprehension, and then watched the video an additional two times.[4] Each video was around two minutes long, and the total training intervention lasted 25 minutes on average. Materials for the two treatment arms were developed specifically for the study and tailored based on insights from focus groups and individual qualitative interviews with individuals similar to our study participants. The interventions were developed in coordination with Development Media International—a non-profit with extensive experience developing media campaigns in West Africa.[5] Participants were randomized at the individual level to one of three distinct training topics, with content varying by treatment arm as follows:

Vaccine Knowledge. The *VaccineKnowledge* training focused on educating participants about the fundamentals of how vaccines build immunity. Specifically, the training covered: (i) how vaccines work—explaining that vaccines do not cure diseases but prepare the body to fight them by training the immune system to recognize and respond to specific threats; (ii) side effects—clarifying that mild side effects, such as feeling unwell for a day or two, are normal and indicate that the body is building immunity; (iii) vaccine schedules and protection—emphasizing that completing the full series of doses is necessary to achieve full protection; and (iv) safety—detailing that vaccines undergo rigorous testing to ensure their safety and effectiveness before being approved.

Misinformation Skills: The *MisinformationSkills* training aimed to help participants identify misinformation using four practical tips. This approach closely resembles the digital media literacy intervention evaluated by Guess et al. (2020), which provided tips for spotting fake news based on Meta's published guidelines. By incorporating examples of misinformation alongside these tips, the training also aligns with inoculation

[4]The structure of the intervention was informed by research in educational psychology which has demonstrated the benefits of repetition (Cepeda et al., 2006) and information retrieval (Roediger and Butler, 2011) on learning outcomes.

[5]See Appendix C for full transcripts of the video interventions

strategies shown to reduce susceptibility to misleading claims (Lewandowsky and Van Der Linden, 2021; Roozenbeek et al., 2022).

The training focused on four features typical of health misinformation in Sierra Leone: (i) emotionally charged language designed to provoke fear or anxiety; (ii) "fake doctors" presenting as credible experts without genuine affiliations; (iii) unverified or untrustworthy sources, prompting participants to question whether the information came from a reliable organization; and (iv) mixed claims combining accurate and false elements, such as "oranges are high in vitamin C and cause cancer." Additionally, participants learned about the potential harms of spreading misinformation, illustrated by an example of someone choosing a harmful alternative treatment over formal healthcare and becoming ill as a result.

Control: To ensure similar levels of engagement and fatigue across study arms, the control training took a similar form to the treatment trainings but focused on a different topic: hand-washing. Focusing on a health behavior also kept the general context—and thus participants' perceptions of the experimenter's preferences—consistent across all study participants.

Our interventions were intentionally broad, designed to build comprehensive knowledge and critical reasoning skills rather than shifting isolated beliefs. While we could have focused narrowly on a few specific beliefs, such as vaccine efficacy or misinformation prevalence, this broader approach allowed us to develop knowledge or skills that may be more transferable to topics not directly covered by the trainings.

3.3 Data Collection

Survey data. All participants completed a brief survey at the time of recruitment. This survey collected information relevant to eligibility—specifically, reading level, smartphone ownership, WhatsApp use, vaccine status, and knowledge of other participants attending the training—as well as basic demographic details. Additionally, because these survey items were pertinent to a separate study, we included questions about formal and informal healthcare utilization.

On arrival at the training session, participants completed an in-person baseline survey immediately prior to the intervention. This survey included basic demographic information—such as age, education, and digital literacy. To provide sample descriptives, the control group was additionally asked to respond to questions about use of and trust in different media sources. After the training session, participants completed a midline survey to assess

immediate impacts on information discernment, learning from and sharing of new information.

One month later, all study participants were invited to complete an in-person endline survey, which measured information discernment, and formal versus informal healthcare utilization over the past month (i.e., since intervention).[6] To measure vaccination uptake, at the end of the endline survey, participants were informed that they could receive the Covid-19 vaccine from a team of government nurses positioned nearby. This mobile vaccine clinic was arranged specifically for the study in collaboration with the Ministry of Health and Sanitation, as the Covid-19 vaccine was no longer widely available in local health clinics at the time. Vaccine uptake was recorded through a survey completed by nurses administering the vaccine.

On-Platform Data.[7] During the baseline survey, participants were asked to provide consent for sharing their WhatsApp conversation data with the research team using a novel data collection tool (Garimella and Chauchard, 2024). This consent was distinct from the general study consent and, due to the potentially sensitive nature of this data, only asked of participants who correctly answered comprehension questions related to scope of data collection. Those who consented were asked to connect to a data collection tool and to select specific chats to share—choosing from a list of their previous individual and group conversations. We attempted to collect data from these selected chats for one month prior to the training (pre-baseline) and for six months following the training (post-baseline).[8] The data collected via the tool includes: i) text messages with names and contact information removed prior to storage, ii) images not deemed 'not safe for work' with facial blurring to ensure anonymization, iii) videos forwarded more than five times,[9], iv) message metadata including timestamps, v) message reactions, and v) de-identified data on participant's network of WhatsApp interactions. All participants could disconnect from the data collection tool at any time.

Pilot studies revealed a high rate of unintended disconnection from the data collection tool in the days and weeks after connection, often due to participants changing phones, phone

[6]The default was to complete this survey at our lab, which was the case for 66% of participants interviewed at endline. When this was not possible, participants were interviewed at a convenient location of their choosing.

[7]As noted above, we are currently human-coding sent and forwarded messages to allow examination of treatment effects on on-platform behaviors.

[8]Due to storage limitations on participants' devices and connectivity constraints, collecting a full month of historical data was not always feasible. Consequently, we gathered as much data as possible within a one-hour window.

[9]As videos could not be reliably de-identified, this restriction, which is in line with WhatsApp's definition of vial content, was designed to reduce the risk of collecting private information.

numbers, or WhatsApp versions. As a result, during the endline survey, we re-consented and reconnected any participants who had not explicitly revoked consent. For those successfully reconnected, we attempted to collect data from one month prior to the endline survey, and aimed to continue data collection for the following six months. As it is possible that the chats available for consent at endline, or the subset of chats consented differ as a result of treatment, we only include reconnected chats that were consented at baseline in our analysis.

3.4 Outcome Measures and Mechanisms

3.4.1 Information Discernment

Our primary discernment outcome was designed with two main aims: (i) to compare whether the interventions produce broad improvements in headline discernment or effects that vary by headline type—for example, true versus false, positive versus negative; (ii) to explore the relationship between gains in vaccine statement discernment and vaccine headline discernment. To meet these aims, we developed a set of statements, some directly related to content covered in the *VaccineKnowledge* intervention (*Core*) and others pertaining to related, but not directly covered, topics (*Adjacent*). Using these statements, we constructed experimental vaccine headlines that recast these statements in a headline-style format, with accompanying images and linguistic elements such as calls to action to mimic real-world media. Each headline directly corresponds to a statement, such that the veracity of the headline logically follows from the veracity of the statement.[10] This design allows us to investigate whether treatment effects on vaccine statement discernment are paralleled by similar gains in vaccine headline discernment for the same content, despite the additional contextual or stylistic elements present in the headlines.

Vaccine Headlines. At both survey horizons, participants saw two sets of headlines: experimental and real. The experimental headlines—our primary discernment measure—were designed to isolate performance across specific features of news items and so are equally balanced across truth (true vs. false), sentiment (positive vs. negative), and relevance (*Core* vs. *Adjacent*). Additionally, to isolate the additive effects of specific misinformation features, all false *Core* headlines lack a source and include: (i) no additional misinformation feature,

[10]For example, the headline "*Taking a vaccine for a sickness gives future protection for that sickness. It's important to protect your health and your community. Take all your vaccine doses and stay safe!*" corresponds to the statement "*If you take a vaccine for a disease, it will protect you from that disease in the future*", and the headline "*Unbelievable Health Miracle! A very big discovery has been made: The marklate can treat you from Corona when you are sick! Quickly go to your nearest health facility for this new treatment!*" corresponds to the statement "*When you are ill, you can get the vaccine to treat that illness.*" The full set of headlines used in the study are accessible here.

(ii) emotional language, (iii) a fake doctor or expert, or (iv) a mix of true and false claims. These experimental headlines were constructed based on the collection and analysis of over 700 headlines shared on WhatsApp and Facebook through local crowdsourcing in Sierra Leone. This process ensured that the experimental headlines reflected locally relevant features and guided the selection of real headlines presented to participants. As these headlines were specifically developed for the study, they eliminate the possibility of prior familiarity influencing responses but may not fully capture the type of news participants encounter in daily life. The real headlines allow us to test whether discernment effects generalize to actual news items.

We elicited participants' beliefs about the likelihood that each headline was true on a scale from 0% to 100% with 25% increments.[11] Headlines were presented in random order within each set (experimental and real). To reduce social desirability bias, we incentivized participants to report their true belief about the accuracy of the headline.[12] Additionally, to avoid treatment effects being driven by shifts in beliefs that the *experimenter* would present false news—potentially more pronounced in the *MisinformationSkills* arm—we informed participants that half of the headlines in each discernment survey module would be false.[13]

To ensure comparability of discernment measures across surveys, experimental headlines were randomly selected (without replacement) for each participant from the same pools of potential headlines at midline and endline. The total pool of experimental headlines comprised four headlines for each possible type (e.g., positive, true, *Core* or positive, false, *Core*) while the total pool of real headlines comprised thirty headlines. At midline, participants evaluated eight experimental headlines, one of each type. At endline, participants evaluated four headlines, one of each type not including *Adjacent* headlines due to time constraints. For real headlines, participants evaluated eight headlines at midline and four at endline, balanced between true and false. Four of these were randomly drawn from the pool of possible headlines—as was also the case for endline—and four corresponded to the headlines participants had the opportunity to share in the sharing exercise.

[11] At both survey horizons, participants completed four examples at the start of the module to ensure scale comprehension. See Appendix Figure B1 for a visualization of the scale provided to participants to aid responses.

[12] Participants were truthfully informed that responses aligning with the actual truth of the headlines would increase their bonus, though the exact bonus calculations were not fully disclosed (following Danz et al. (2022))

[13] Prior to fixing beliefs, we see differences in expectations across treatment groups. Participants were asked: "We have put together ten Whatsapp messages on health topics. For each message, I will ask you if you think that all the information in this message is TRUE. Out of the ten messages I am going to show you, how many do you think will contain false information?" On average, participants in the control group expected 3.61 false headlines. In the *MisinformationSkills* arm, participants expected significantly more false messages (0.94, SE: 0.13), while there was no statistically significant difference in the *VaccineKnowledge* arm (-0.13, SE: 0.13).

Additionally, we included two types of experimental headlines to assess whether the *MisinformationSkills* training leads participants to distrust obviously true information presented without a verified source or, conversely, to trust headlines with a false but credible-looking source. For the former, participants were shown one of two headlines: "*Smoking is Dangerous! Smoking is bad for your health. Quit smoking now!*" or "*Vaccines in Sierra Leone! Children can take vaccines in Sierra Leone!*"[14]

Vaccine Statements. We developed vaccine-related statements both to assess participants' ability to discern vaccine information not in a headline format and to inform the content of the experimental headlines. These statements draw on material covered in the *VaccineKnowledge* training (*Core*) and topics not explicitly addressed in the training (*Adjacent*) but logically related.[15] For example, "*Adults can have side effects after they take a vaccine*" is a *Core* statement as this information was directly covered by the training, while "*Children can sometimes be paralyzed after taking the measles vaccine*" is an *Adjacent* statement as this was not directly addressed by the training. This categorization allows us to evaluate discernment for directly taught material and the transfer of knowledge to related but unaddressed topics. Statements varied in both sentiment (positive vs. negative) and truth (true vs. false), but they were not evenly distributed across these categories, as our goal was not to compare statement discernment on these attributes.

Similar to headlines, participants rated the likelihood that each statement was true on a scale from from 0% to 100% in 25% increments.[16] As with the elicitation of headline discernment, participants were truthfully informed that responses aligning with the factual status of the statements would increase their bonus, and the statements were presented in random order. Additionally, the order of the vaccine statement module and experimental vaccine headline module was randomized.

3.4.2 Information Behaviors

To compare how the interventions affect broader information-processing and dissemination behaviors, we measure two distinct outcomes: participants' ability to learn from new vaccine-related information and their propensity to share vaccine content with others.

[14]Control group mean responses suggests this information is well-known to be true, with 92% and 93% of participants responding that the headlines were 100% likely to be true, respectively.

[15]To provide sample descriptives, we included an additional statement about vaccines causing infertility in this module; however, since this misconception was uncommon in our target population, it is not excluded in our analyses in line with our pre-registration.

[16]As for headlines, at both survey horizons, participants completed four examples at the start of the module to ensure scale comprehension. See Appendix Figure B1 for a visualization of the scale provided to participants to aid responses.

Learning from Information. To measure participants' ability to learn from new vaccine-related information, we designed a survey exercise that mimics a scenario in which an individual wants to form accurate beliefs about a topic after encountering information that may be true or false. For instance, this might reflect a parent searching online for information about a new medicine or someone deciding whether to take the Covid-19 vaccine after receiving forwarded messages from friends or family on WhatsApp.

We asked participants two incentivized questions about the safety and efficacy of the HPV vaccine:[17] one required them to estimate the percentage of young girls who would die as a result of taking the vaccine, and the other asked them to estimate the percentage of cervical cancers the vaccine could prevent. To isolate treatment effects on learning from effects on prior beliefs, we randomly assigned participants to one of two groups. Half the sample responded without receiving any additional information. The other half responded after reviewing a curated set of real headlines about the HPV vaccine, including both accurate claims, such as the vaccine's high efficacy in preventing cervical cancer, and inaccurate claims, such as false reports of fatal side effects. Participants were not prompted to evaluate the veracity of the headlines, but were told that the "headlines might be true or false" and that they would be asked questions about the headlines and could earn money based on the accuracy of their responses.

Sharing Discernment: To examine effects on participants' propensity to share both true and false information, we designed an exercise mirroring the incentives people face when sharing news on WhatsApp. Each participant was given the option to share up to four news items about the malaria vaccine—two false and two true—with other study participants recruited from their community. To prevent false information from spreading beyond the lab—where we could not debrief recipients—shared items were restricted to a WhatsApp group accessible only to participants in the lab. However, to mirror the reputational incentives of real-world information sharing, shared information was accompanied by the participant's WhatsApp handle. Participants expected, on average, 6.7 out of 10 people from their community viewing the group to recognize their WhatsApp handle, and only 7% reported that zero people would recognize it. This suggests that reputational costs to sharing information within this lab exercise were present for a large majority of participants. Participants were aware of these features of the task when making sharing decisions. Our measure of sharing discernment is the difference between the percentage of true and false news items shared.

[17]The HPV vaccine was made available to adolescents in Sierra Leone in October 2022, seven months before the start of the study.

3.4.3 Downstream Health Outcomes

Finally, in the endline survey, we evaluate the impacts of the treatments on two policy-relevant health outcomes: vaccine uptake and formal versus informal healthcare utilization.

Vaccine Take-up: After completing the survey component, participants were given the opportunity to receive the Covid-19 vaccine at a location next to the lab from a team of nurses deployed by the Ministry of Health and Sanitation specifically for the study. At the time of the study, Covid-19 vaccines were no longer available in clinics in Freetown. To verify that participants had not obtained the vaccine privately between the training and endline, we asked whether they had received the Covid-19 vaccine in the past month and, when possible, verified this with their vaccination card.[18]

Healthcare Utilization Index: To assess changes in health-seeking behaviors, particularly shifts from informal to formal providers, we collected self-reported healthcare practices over the past month in the endline survey. Positively coded behaviors include consulting formal providers, such as doctors, nurses, or clinics, and purchasing medications from pharmacists. Negatively coded behaviors include obtaining drugs from informal sources, such as peddlers, market vendors, or online sellers. As these components are likely to be highly correlated, we construct the index following Anderson (2008), using weights based on the inverse covariance matrix of residualized, standardized outcomes. This approach down-weights highly correlated outcomes, prioritizing those that capture independent variation in healthcare behaviors.

3.5 Balance and Attrition

The randomization was successful: Appendix Table A2 shows that out of eleven covariates compared across arms, one—whether a participant had below median income—is statistically significantly different across arms ($p = 0.01$). Reassuringly, we see no statistically significant differences in likely correlates of income that could affect treatment effects, including reading level, education, digital literacy, and WhatsApp use.

Appendix Table A3 shows that attrition is relatively low at 10% overall, with similar endline completion rates across study arms. The likelihood of being interviewed in the lab versus at a convenient location at endline (66% for the full sample) is also similar across arms. In the post-attrition sample, several baseline variables differ across treatments—specifically, whether the participant is a woman ($p = 0.05$), has below median income ($p = 0.01$), and has

[18]Fewer than one percent of participants (N=10) report obtaining the vaccine, but none could be verified through vaccination cards, suggesting these reports were due to social desirability bias.

taken the Covid-19 vaccine ($p = 0.04$). To test whether these differences reflect systematic attrition, we estimate a regression of endline completion on each baseline variable and its interactions with treatment assignment. The resulting F-test ($p = 0.22$) provides no evidence of differential attrition across treatment arms.[19]

4 Empirical Strategy

For discernment outcomes, we estimate average treatment effects at the headline or statement level using the following OLS regression model:

$$\begin{aligned} y_{ih} = \beta_0 + \beta_1 \text{VaccineKnowledge}_i + \beta_2 \text{MisinformationSkills}_i \\ + \phi X_i' + \lambda_h + \gamma_c + \epsilon_{ih} \end{aligned}$$

where y_{ih} is the discernment score of participant i for headline, or statement, h. This score is constructed by reverse-coding responses to false items so that a score of 100 reflects perfect discernment—that is, complete alignment with the factual status of the headline or statement—and standardizing each score using the control group mean and standard deviation at the headline level. $\textit{VaccineKnowledge}_i$ and $\textit{MisinformationSkills}_i$ are binary indicators for treatment assignment to the *VaccineKnowledge* or *MisinformationSkills* group, respectively. Regressions control for community fixed effects (γ_c), item fixed effects (λ_h), and baseline covariates (X_i'). Baseline covariates include variables selected via post-double-selection LASSO (Belloni et al., 2014) from a pre-specified set of potential covariates, as well as survey design variables, such as module order and whether the participant was randomly assigned to justify their answer after the experimental headline or sharing modules. Standard errors are clustered at the participant level (the unit of randomization). To check robustness to alternative specifications and for non-discernment outcomes, we estimate participant-level regressions of an equivalent form.

For learning from information, since we split the sample into two groups–one providing data to estimate priors (without headline exposure) and the other to estimate posteriors (after headline exposure)—we pre-registered a difference-in-differences regression approach to estimating average treatment effects of the following form:

[19] Appendix Table A7 presents inverse-probability weight (IPW) adjusted results for our main outcomes, showing that our results are robust to this adjustment.

$$
\begin{aligned}
Y_{it} = \beta_0 &+ \beta_1 \text{Post}_{it} + \beta_2 \text{MisinfoTreatment}_i + \beta_3 \text{VaccineTreatment}_i \\
&+ \beta_4 (\text{MisinfoTreatment}_i \times \text{Post}_{it}) \\
&+ \beta_5 (\text{VaccineTreatment}_i \times \text{Post}_{it}) \\
&+ \phi X'_{it} + \gamma_c + \epsilon_{it}
\end{aligned}
$$

where y_{it} is the outcome for participant i and time t, with $t = 0$ indicating priors (without headline exposure) and $t = 1$ indicating posteriors (after headline exposure). $Post_{it}$ is a dummy indicating post-headline measurement. All other variables are as previously defined.

To account for multiple hypothesis testing within each pre-registered outcome family, we report FDR-adjusted q-values using the two-stage procedure of Benjamini et al. (2006).

Pre-Analysis Plan While we adhere to our pre-analysis plan (PAP) in most respects, we deviate in two key ways. First, the PAP specified regressions including all interactions between treatment arms and headline characteristics (e.g., truth, sentiment, and relevance). These regressions, though comprehensive, result in numerous coefficients that are challenging to interpret. To address this, we focus on subsets of headline, such as true headlines or other pre-specified outcome families, as our primary analyses.

Second, while we initially pre-specified multiple hypotheses adjustment within six families of outcomes, we deviate by presenting results for real vaccine headline discernment, the sharing discernment index, and learning from information in separate tables. Adjusting within this pre-specified family does not alter the significance patterns of our results (Appendix Table A8). Therefore, we present these results across separate main tables to facilitate more meaningful comparisons.

5 Results

5.1 Effects on Information Discernment

Table 1 presents treatment effects on discernment of experimental headlines in Column 1. The control group mean is 77.75% percent, meaning that, on average, participants rated true headlines as 77.75% likely to be true and false headlines (after reversing their responses) as 77.75% likely to be false. This indicates an ability to distinguish true from false headlines that exceeds random guessing (50 percent) but falls short of perfect discernment (100 percent), with substantial variation across individuals (SD = 34.38). Indeed, only 16 percent of

the control group achieve a perfect score across the set of eight headlines, suggesting scope for improvement among a majority of participants. Responses reveal an asymmetry: participants rate true headlines as 88.97% likely to be true but false headlines only 66.53% likely to be false, suggesting greater susceptibility to false claims.

Both treatments lead to small but statistically significant improvements in headline discernment. The *VaccineKnowledge* arm increases discernment by 0.09 SD ($q < 0.01$), while the *MisinformationSkills* arm yields a larger improvement of 0.18 SD. ($q < 0.01$). These effects differ significantly ($q < 0.01$). For context, the improvement from the *VaccineKnowledge* arm is comparable to the difference in discernment between control participants with a secondary education and those with a university degree, while the improvement from the *MisinformationSkills* arm exceeds this educational gradient.

Although real headlines differ from experimental ones—they do not include false-positive headlines and were partially influenced by sharing decisions—Appendix Table A4, Column 4, shows treatment effects of similar magnitudes. However, unlike for experimental headlines, these effects do not significantly differ across treatments. One possible explanation is the difference in headline type across modules. When we exclude false-positive headlines from the experimental set—aligning them with real headlines in terms of veracity and sentiment (Appendix Table A4, Column 1)—we also find no significant difference between treatments (p =0.94). This suggests that the *VaccineKnowledge* arm may operate differently on false-positive headlines, a possibility we explore further in Section 5.1.1.

To assess the robustness of our findings across different outcome definitions and to facilitate comparisons with other studies, Appendix Table A5 presents treatment effects on multiple specifications. These include: (1) our preferred specification-the standardized outcome variable analyzed at the headline level (Column 1, identical to Table 1, Column 1); (2) the unstandardized version of the outcome variable (Column 2); (3) the proportion of news items correctly identified, where false news rated 0% likely to be true and true news rated 100% likely to be true (Column 3); and (4) an alternative correctness threshold, where false news rated 0% or 25% likely to be true and true news rated 75% or 100% likely to be true are considered correct (Column 4). Columns 5 to 8 replicate these specifications, but with analysis at the participant level-i.e., averaging across headlines for each individual-rather than at the headline level. Panel A reports results for experimental headlines, while Panel B presents results for real-world headlines, which follow similar patterns to those for experimental headlines.

Our results remain robust to alternative definitions of the outcome variable. Both treatments increase the likelihood of correctly identifying headline truthfulness by 8.0 p.p. (p <0.01) (Column 3). For comparison, Ali and Qazi (2023) report a 3.3 percentage point in-

crease on a similar outcome measure following a video-based misinformation intervention in urban Pakistan. A recent mega-study comparing nine interventions to reduce belief in false headlines found that most interventions had statistically significant effects, though none exceeded 0.10 on a six-point Likert scale (Fazio et al., 2024). Converting this 0.10 effect to our probability-based scale ($0.10 \times (100/6)$) yields an implied treatment effect of approximately 1.7 p.p., aligning with the 2.47 p.p. (p <0.01) and 6.21 p.p. (p <0.01) effects observed at the participant level for *VaccineKnowledge* and *MisinformationSkills* (Column 6), respectively. Together, these findings suggest that our treatments produce effects comparable to, or larger than, many existing interventions.

5.1.1 Information Discernment Mechanisms

We identify four pathways through which the treatments may influence headline discernment, either positively or negatively.

Turning to potential mechanisms, we posit four pathways through which the treatments might influence headline discernment, either positively or negatively. First, knowledge: The *VaccineKnowledge* treatment strengthens participants' foundational vaccine knowledge, enabling them to evaluate headline content against established facts. Second, skills: The *MisinformationSkills* treatment provides heuristics to identify misleading headlines, enhancing discernment without requiring extensive subject knowledge. Third, skepticism: Misinformation interventions may increase generalized distrust, which could improve the ability to recognize false content but also undermine confidence in accurate information (Hoes et al., 2024). Fourth, overoptimism: Greater trust in vaccines, fostered by the *VaccineKnowledge* treatment, may bias participants toward accepting positive vaccine-related headlines, even when false, thereby reducing discernment. This mechanism reflects a tendency for individuals to judge the accuracy of new information based on its consistency with preexisting beliefs, as observed in evaluations of politically aligned news (Thaler, 2024).

To examine these mechanisms, we analyze performance across different *types* of headlines (Table 1, Columns 2 to 7), and, for the knowledge mechanism, vaccine-related statement discernment (Table 1, Columns 8 and 9).

Knowledge. To assess the role of vaccine-related beliefs in shaping headline discernment, we analyze treatment effects on statement discernment and their alignment with effects on experimental headline discernment. Table 1, Column 8, shows that the *VaccineKnowledge* treatment significantly improves discernment for *Core* vaccine statements (0.26 SD, $q <$ 0.01); however, this effect does not extend to *Adjacent* statements (Column 9)—content not directly addressed in the training (-0.00 SD, q =0.33), indicating limited transfer beyond

the material directly covered . By contrast, the *MisinformationSkills* arm has no statistically significant effect on discernment for *Core* statements (-0.01 SD, $q = 0.27$) but leads to a significant improvement for *Adjacent* statements (0.10 SD, $q < 0.01$). These effects suggest that the training's emphasis on identifying emotion-driven claims—more common in *Adjacent* than *Core* statements—may drive these effects.

Comparing these statement discernment patterns with experimental headline discernment results (Table 1, Columns 6 and 7) reveals a similar pattern for *VaccineKnowledge*: discernment improves for *Core* vaccine headlines (0.16 SD, $q < 0.01$) but does not show statistically significant gains for *Adjacent* headlines (0.02 SD, q =0.19). By contrast, the *MisinformationSkills* arm improves discernment for headlines on both *Core* (0.17 SD, q <0.01) and *Adjacent* topics (0.18 SD, q <0.01), despite having no measurable effect on *Core* vaccine statement discernment. These results suggest that improvements in the *VaccineKnowledge* arm are closely tied to knowledge aquisition, whereas the *MisinformationSkills* treatment appears to rely on broader heuristics for identifying false information, independent of specific content knowledge.

Skills. We begin by examining whether discernment gains in the *MisinformationSkills* arm are primarily driven by source evaluation or the use of additional heuristics, analyzing performance across different headlines types. Experimental true headlines lack misinformation features and so cannot shed light on the heuristics participants used. *Core* false headlines always lacked a source and, at random, either: (i) contained no additional feature, or included (ii) emotional language, (iii) a fake doctor or expert, or (iv) a mix of true and false claims. *Adjacent* false headlines also lacked a source but included misinformation features not explicitly covered in the training—such as spelling mistakes, formatting errors, hashtagged terms, and colloquial abbreviations (e.g., "vax" for vaccine, or "u" for you). Since all false headlines lacked a source, we cannot isolate the effect of source evaluation from the effects of other heuristic cues. However, comparing performance across different headline types allows us to assess whether discernment improves when multiple misinformation features are present (additive complementarity) or whether source evaluation substitutes for other heuristics in driving discernment gains.

Focusing on the *MisinformationSkills* arm, we find no statistically significant difference in treatment effects for *Core* false headlines without a source compared to those with no source plus an additional misinformation feature ($p = 0.49$, Table A6, Col. 2). Similarly, we cannot reject the equality of treatment effects for *Adjacent* false headlines—those without a source but with an untrained feature of misinformation—and *Core* false headlines with taught features ($p = 0.65$, Appendix Table A6, Col. 3). These findings suggest that

source evaluation is a primary driver of discernment gains. Yet, two results indicate that source evaluation alone does not fully explain discernment gains. First, Appendix Table A4, Column 3, shows that the *MisinformationSkills* arm improves discernment for real false headlines, which—unlike the experimental headlines—contain a non-reliable source rather than no source. This suggests that discernment gains persist even when source credibility is harder to evaluate. Second, participants in the *MisinformationSkills* arm exhibit discernment improvements for *Adjacent* false statements (0.10 SD, $q < 0.01$), which often include emotionally evocative claims, pointing to the use of emotional language heuristics.

Together, these findings suggests that while multiple heuristics may contribute to discernment gains, source evaluation appears to be the dominant factor, with no evidence of additive complementarity between source evaluation and other heuristics.

Skepticism. Table 1, Column 2, presents treatment effects on discernment of true experimental headlines. We find no evidence of an indiscriminate increase in skepticism—participants in both arms show improvements in discernment for true headlines: 0.13 SD ($q < 0.01$) for *VaccineKnowledge* and 0.08 SD ($q <0.01$) for *MisinformationSkills*.

Overoptimism. To test whether participants in the *VaccineKnowledge* arm exhibit weaker discernment for positive headlines, we begin by analyzing treatment effects on positive experimental headlines (Table 1, Column 4). The *VaccineKnowledge* arm exhibits performance gains relative to the control group (0.06 SD, $q = 0.03$) but underperforms compared to *MisinformationSkills* (0.19 SD, $q < 0.01$), with the difference between the two effects statistically significant ($q <0.01$). In contrast, for negative headlines, both treatments yield statistically significant discernment gains, and we cannot reject equality across treatment arms ($q =0.10$).

These analyses provide suggestive evidence on the distinct mechanisms through which the treatments influence headline discernment. The *VaccineKnowledge* treatment improves discernment by enhancing participants' foundational knowledge, enabling them to evaluate headlines aligned with the training content. Its effects are narrower, with limited transfer to topics beyond the training, but participants show substantial gains in understanding *Core* vaccine statements, underscoring the intervention's efficacy in strengthening key vaccine-related knowledge.

By contrast, the *MisinformationSkills* treatment improves discernment across a broader range of headline types, with evidence pointing to a major role for source evaluation, though the use of other heuristics cannot be ruled out. Together, these findings suggest that while both treatments improve discernment, they do so through distinct pathways: *VaccineKnowl-*

edge builds specific content knowledge, while *MisinformationSkills* fosters transferable skills for identifying misleading claims. Additionally, we find no evidence of potential adverse effects on discernment—overoptimism in the VaccineKnowledge arm and skepticism in the MisinformationSkills arm—materializing in this context.

5.1.2 Information Discernment Persistence

Table 2 examines the persistence of treatment effects on discernment across survey horizons, presenting midline results in Panel A and endline results in Panel B, along with tests of equality of effects across time points. To ensure comparability, we restrict the analysis to headlines randomly drawn, at the participant level, from the same pool of potential headlines. For experimental headlines (Columns 1 to 3), we include only *Core* headlines, as *Adjacent* headlines were not assessed at endline. For real headlines (Columns 4 to 6), we include only the four headlines not used in the midline sharing exercise at midline, which were drawn from the same pool as the endline real headlines.

The results reveal persistent treatment effects on discernment. For experimental *Core* headlines, both treatments show lasting effects. *VaccineKnowledge* improves discernment by 0.21 SD ($p < 0.01$), directionally higher but not significantly different from the midline result ($q = 0.10$). *MisinformationSkills* improves discernment by 0.22 SD ($p < 0.01$), similarly not statistically distinguishable from midline ($q = 0.16$). A similar pattern emerges for real headlines: *VaccineKnowledge* increases discernment by 0.13 SD ($p < 0.01$) and *MisinformationSkills* increase discernment by 0.20 SD ($p < 0.01$).

5.2 Effects on Information Behaviors

Building on the treatment effects observed for headline discernment, we next investigate downstream information behaviors, focusing on learning and sharing.

5.2.1 Learning from Information

To assess differences in learning from new headlines across treatment arms, participants provided incentivized estimates for two HPV vaccine-related questions—one on vaccine safety and the other on efficacy—either without prior exposure to related headlines or after viewing a set of HPV vaccine-related headlines. Table 3 presents results for the safety (Columns 1 and 2) and the efficacy question (Columns 3 and 4). Odd-numbered columns show treatment effects on participants' prior and belief updating (the difference between priors and posteriors). Even-numbered columns focus on posterior beliefs among participants exposed to headlines before answering questions, capturing treatment effects on learning from information.

Starting with beliefs about vaccine safety, Column 1 of Table 3 shows that prior to headline exposure, participants in the *VaccineKnowledge* arm report slightly more accurate priors than the control group (-0.64 p.p. $p = 0.71$), though the difference is not statistically significant. Participants in the *MisinformationSkills* arm exhibit slightly less accurate priors (3.11 p.p., $p = 0.09$), with the difference between treatment arms marginally significant ($p = 0.06$). After headline exposure, control group participants report less accurate beliefs, with accuracy declining by 5.72 p.p. ($p < 0.01$). In contrast, participants in the *VaccineKnowledge* arm do not exhibit this decline–the interaction term of -5.56 ($p = 0.04$) effectively offsets the control group's negative update. The *MisinformationSkills* arm shows no statistically significant difference in updating patterns from the other two groups (-3.29, $p = 0.25$). As a result, posterior accuracy for vaccine safety remains significantly higher in the *VaccineKnowledge* arm than in both the control and *MisinformationSkills* arms, reflecting a combination of more accurate priors and greater resistance to misinformation.

Turning to beliefs about HPV vaccine efficacy, Column 3 of Table 3) shows that, prior to headline exposure, participants in the *VaccineKnowledge* arm report significantly more accurate priors than those in the *MisinformationSkills* arm (-2.37 p.p. vs. 3.20 p.p., both $p < 0.05$). The difference between treatment arms is statistically significant ($p < 0.01$). However, headline exposure induces minimal belief updating—less than 1 p.p. on average, with no significant differences across arms. Consequently, the initial advantage in accuracy for *VaccineKnowledge* participants persists, but this reflects baseline knowledge differences rather than differential learning from new information.

Together, these results show that the *VaccineKnowledge* arm starts with more accurate priors and is better protected against false headlines. We cannot rule out similar protection in the *MisinformationSkills* arm. Despite discernment gains, neither treatment improves learning from new information —beliefs do not update in a more accurate direction. This suggests that beyond discernment, additional factors are necessary for true information to meaningfully improve belief accuracy.

5.2.2 Sharing

Engagement with the exercise was high across all groups, with 95% of participants sharing at least one headline. Table 4 reports treatment effects on sharing discernment (Column 1), the percentage of true (Column 2) and false headlines shared (Column 3), the likelihood to share any headline (Column 4) and the number of headlines shared (Column 5). The *MisinformationSkills* treatment improves sharing discernment by 6 p.p. ($p = 0.04$), driven by a statistically significant 7 p.p. ($p = 0.01$) reduction in the percentage of false headlines shared. The *VaccineKnowledge* treatment shows no statistically significant effect on sharing

discernment (5 p.p., $p = 0.11$) or on the percentage of true and false headlines shared. However, we cannot reject equality of treatment effects across the two arms on sharing discernment (p =0.71) or the percentage of false headlines shared (p =0.27). Crucially, both treatments have precisely estimated null effects on the percentage of true headlines shared, indicating that reductions in false headline sharing do not come at the expense of sharing accurate content. Taken together, these results suggest that both treatments enhance sharing discernment to a similar degree, consistent with the comparable treatment effects observed for headline discernment.

5.3 Effects on Downstream Health Outcomes

The persistent effects on discernment and vaccine knowledge allow us to examine whether these gains translate into downstream health behaviors relevant to policy-makers, specifically Covid-19 vaccine uptake and formal healthcare utilization. Table 5 , Column 1, reports treatment effects on Covid-19 vaccination uptake, showing no statistically significant differences compared to control group or across treatment arms. The estimated effects are 0.02 percentage points (q =1.00) for *VaccineKnowledge* and 0.04 (q =1.00) for *MisinformationSkills*. Vaccination uptake in the control group, at 27%, was higher than expected, which reduced statistical power to detect treatment effects, as power calculations assumed a control group mean of 10%. We can rule out increases in vaccination uptake larger than 8 percentage points for *VaccineKnowledge* and 10 percentage points for *MisinformationSkills*. Given that the headline discernment effects are of comparable magnitude, the absence of a significant effect on vaccination uptake is perhaps not surprising.

Table 5, Column 2, presents treatment effects on the Healthcare Utilization Index, which captures engagement with formal versus informal health services and medication purchases. The results indicate no significant effects on the overall index, with estimates of 0.02 standard deviations (q =1.00) for (*VaccineKnowledge* and -0.01 standard deviations (q =1.00) for *MisinformationSkills*. Similarly, there are no significant effects on most subcomponents. These findings suggest that improvements in knowledge and discernment do not consistently translate into broader healthcare behavior changes, a point explored further in the Discussion.

6 Conclusion

Public health initiatives frequently rely on targeted education to correct health misconceptions. However, as information increasingly circulates online, equipping individuals with

skills to evaluate accuracy—an strategy commonly used to counter misinformation more broadly—may serve as an effective alternative. To assess this approach, we conducted a randomized controlled trial in Freetown, Sierra Leone, examining the effects of two interventions on vaccine information discernment, as well as downstream information and health behaviors.

We find that both interventions improve vaccine headline discernment, with effects persisting one month later. However, their impact differs in scope. The *VaccineKnowledge* intervention enhances discernment only for headline content directly related to trained topics, with no spillover to related but unaddressed topics. By contrast, the *MisinformationSkills* intervention leads to broader improvements across both topic and sentiment. Effects also differ by information format: the *VaccineKnowledge* arm leads to lasting improvements in factual statement discernment when information is read aloud by surveyors—potentially mimicking how individuals evaluate information in conversation—while *MisinformationSkills*' effects are primarily limited to headline evaluation. These differences suggest that the optimal policy approach depends on the specific misinformation risks within a given domain.

Despite discernment effects comparable to those of other interventions, evidence for downstream behavior impacts is mixed. Participants in the *VaccineKnowledge* arm are less susceptible to false claims when presented with new information, yet neither intervention improves belief accuracy relative to those receiving no additional information. Both treatments enhance sharing discernment, though the effect for *VaccineKnowledge* is not statistically significant. Additionally, we find no evidence that participants shift away from informal healthcare providers or increase vaccine uptake. Given these mixed downstream effects, we outline three potential directions for future research.

First, to our knowledge, no pior work examines how discernment gains affect individuals' ability to *learn* from new information. While discernment improves, participants exposed to both true and false information do not form more accurate beliefs. Future research could explore the generalizability and mechanisms underlying this finding. Misinformation interventions often work by making specific features of information more salient (Pennycook and Rand, 2022), raising the question of whether individuals fail to apply discernment when accuracy is not explicitly emphasized. Alternatively, discernment may be necessary but not sufficient for learning, as other barriers—such as information complexity—could limit belief updating.

Second, the policy effectiveness of these interventions depends not only on their impact but also on their demand. In highly politicized settings—such as vaccination in the United States—delivering domain knowledge outside formal education settings may be particularly challenging. At the same time, individuals may be more receptive to learning about widely

discussed or personally relevant health topics than to interventions focused on misinformation detection, a skill in which people often exhibit overconfidence (Lyons et al., 2021; Taavila, Amara, Draginić, Jung, Li, and Sloksnath, Taavila et al.). Future research could explore how demand for different misinformation interventions varies by framing and delivery, shedding light on potential barriers to scaling these approaches effectively.

Finally, recent research has developed theoretical frameworks to explain the spread and discernment of political information, where sharing is often driven by partisan signaling (Guriev et al., 2023) and discernment is shaped by ideological alignment (Thaler, 2024). However, less is known about the motivations driving the sharing of health and other non-political information. Future research could examine whether distinct mechanisms, such as altruism in sharing accurate health information or expertise signaling in engaging with complex medical topics, influence behavior.

Table 1: Primary Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Experimental Headline Discernment for:						Statement Discernment for:	
	Experimental Headline Discernment	True Headlines	False Headlines	Positive Headlines	Negative Headlines	Core Topic Headlines	Adjacent Topic Headlines	Core Topics	Adjacent Topics
VaccineKnowledge	0.09***	0.13***	0.05	0.06*	0.12***	0.16***	0.02	0.26***	-0.00
(V)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
	[<0.01]	[<0.01]	[0.08]	[0.03]	[<0.01]	[<0.01]	[0.19]	[<0.01]	[0.33]
MisinformationSkills	0.18***	0.08**	0.27***	0.19***	0.16***	0.17***	0.18***	-0.01	0.10***
(M)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
	[<0.01]	[0.01]	[<0.01]	[<0.01]	[<0.01]	[<0.01]	[<0.01]	[0.27]	[<0.01]
Control Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Unstandardized Control Mean	77.75	88.97	66.53	77.56	77.94	76.11	79.39	67.45	54.16
Unstandardized Control SD	34.38	24.85	38.67	35.27	33.48	35.15	33.53	39.34	42.46
N	14520	7260	7260	7260	7260	7260	7260	14513	9071
Number of Participants	1815	1815	1815	1815	1815	1815	1815	1815	1815
Postestimation tests:									
p-value: V=M	<0.01	0.10	<0.01	<0.01	0.26	0.80	<0.01	<0.01	<0.01
q-value: V=M	[<0.01]	[0.04]	[<0.01]	[<0.01]	[0.10]	[0.27]	[<0.01]	[<0.01]	[<0.01]

Each column presents treatment effect estimates relative to control with observations at the headline (Columns 1 to 7) or statement (Columns 8 and 9) level. The dependent variables are indicated in the column title. For Columns 1 to 7, the dependent variable is the experimental headline discernment, measured on a 0 to 100 scale. This metric captures how accurately participants assessed the probability of a headline being entirely true. If the headline was actually true, their score corresponds directly to their assessment. Conversely, if the headline was false, their score is calculated as 100 minus their assessment. For column 8 to 9, the dependent variable is the vaccine beliefs index, measured on a 0 to 100 scale, with statement adjacent 5 excluded as pre-specified. Scores were calculated in the same manner as described for the headline discernment metric. Standard errors clustered at the individual level are in parentheses and FDR-adjusted p-values following the two-stage approach in Benjamini, Krieger, and Yekutieli (2006) are in square brackets. The adjustment is done within two families: family 1 includes column 1 only, family 2 includes columns 2 to 9. All regressions control for community and item fixed-effects as well as for module order. Additional controls selected via double-LASSO selection are as follows: the respondent's gender (Columns 2, 8), whether the respondent is engaged in post-secondary education (Columns 3, 9), whether the respondent knows how to report someone on WA (Columns 1, 3, 5, 6, 7, 8), the respondent's reading level (Columns 1, 3, 4, 5, 6, 7, 8, 9), whether the respondent passed an initial attention check (Columns 1, 3, 7, 8, 9), whether the respondent refused to share their monthly income (Columns 1, 2, 3, 4, 5, 6, 7, 8, 9).

* p<0.10, ** p<0.05, *** p<0.01

Table 2: Persistence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Experimental Headline Discernment for:			Real Headline Discernment for:			
	Core Headlines	Core True Headlines	Core False Headlines	Non-Sharing Headlines	Non-Sharing True Headlines	Non-Sharing False Headlines	Statement Discernment
Panel A: Midline							
VaccineKnowledge	0.16***	0.26***	0.07	0.11***	0.13***	0.09**	0.16***
(V)	(0.03)	(0.04)	(0.05)	(0.03)	(0.04)	(0.05)	(0.02)
MisinformationSkills	0.17***	0.07	0.27***	0.10***	-0.03	0.23***	0.03
(M)	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.02)
Control Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Unstandardized Control Mean	76.11	89.14	63.08	75.75	84.28	67.21	62.34
Unstandardized Control SD	35.15	24.08	39.40	34.80	29.50	37.51	41.08
N	7260	3630	3630	7256	3628	3628	23584
Number of Participants	1815	1815	1815	1814	1814	1814	1815
Panel B: Endline							
VaccineKnowledge	0.21***	0.51***	-0.09*	0.13***	0.27***	-0.02	0.10***
(V)	(0.03)	(0.04)	(0.05)	(0.03)	(0.05)	(0.05)	(0.02)
MisinformationSkills	0.22***	0.24***	0.20***	0.20***	0.27***	0.12**	0.03
(M)	(0.03)	(0.05)	(0.05)	(0.03)	(0.05)	(0.05)	(0.02)
Control Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Unstandardized Control Mean	67.86	69.04	66.69	65.89	56.77	75.00	59.00
Unstandardized Control SD	38.30	38.46	38.12	37.18	40.64	30.79	40.35
N	6571	3285	3286	6570	3284	3286	21369
Number of Participants	1643	1643	1643	1643	1643	1643	1644
Postestimation tests:							
p-value: $V_{Midline} = V_{Endline}$	0.17	<0.01	0.03	0.85	0.02	0.04	0.02
q-value: $V_{Midline} = V_{Endline}$	[0.10]	[<0.01]	[0.04]	[0.28]	[0.04]	[0.05]	[0.04]
p-value: $M_{Midline} = M_{Endline}$	0.43	<0.01	0.10	0.07	<0.01	0.02	0.42
q-value: $M_{Midline} = M_{Endline}$	[0.16]	[0.03]	[0.08]	[0.06]	[<0.01]	[0.04]	[0.16]

Each column presents treatment effect estimates relative to control with observations at the headline (Columns 1 to 6) or statement (Columns 7) level. The dependent variables are indicated in the column title. For Columns 1 to 3, the dependent variable is the experimental headline discernment, measured on a 0 to 100 scale. This metric captures how accurately participants assessed the probability of a headline being entirely true. If the headline was actually true, their score corresponds directly to their assessment. Conversely, if the headline was false, their score is calculated as 100 minus their assessment. For Columns 4 to 6, the dependent variable is the real headline discernment, measured on a 0 to 100 scale. Scores were calculated in the same manner as described for the experimental headline discernment metric. For column 7, the dependent variable is the vaccine beliefs index, measured on a 0 to 100 scale. Scores were calculated in the same manner as described for the headline discernment metric. Panel A presents result at Midline for headlines which were shown both at midline and endline, while Panel B presents result at Endline. All statements except statement adjacent 5 are included in both panels as pre-specified. Standard errors clustered at the individual level are in parentheses. All regressions control for community and item fixed-effects as well as for module order. Following a double-LASSO selection, we also control for the respondent's gender (Column A7), whether the respondent is engaged in post-secondary education (Columns A3, B1, B3, B4, B6, B7), whether the respondent knows how to report someone on WA (Columns A1, A3, A4, A6, A7, B1, B3, B6, B7), whether the respondent uses google at least once a day (Columns B4, B6), the respondent's reading level (Columns A1, A3, A4, A6, A7, B1, B3, B4, B6, B7), whether the respondent passed an initial attention check (Columns A7, B7), whether the respondent's baseline/midline survey was lost due to an SCTO error (Columns B1, B3), whether the respondent refused to share their monthly income (Columns A1, A2, A3, A4, A5, A6, A7, B1, B2, B3, B4, B5, B6, B7).

* p<0.10, ** p<0.05, *** p<0.01

Table 3: Learning Index, Midline

	(1)	(2)	(3)	(4)
	Vaccine Safety (Q1) Accuracy		Vaccine Efficacy (Q2) Accuracy	
	For All Participants	For Headline Viewers	For All Participants	For Headline Viewers
Priors				
VaccineKnowledge (V)	-0.64		-2.37**	
	(1.73)		(1.21)	
MisinformationSkills (M)	3.11*		3.20**	
	(1.83)		(1.48)	
Update				
Update	5.72***		-0.08	
	(1.91)		(1.37)	
Update X V	-5.56**		0.80	
	(2.64)		(1.82)	
Update X M	-3.29		-1.11	
	(2.84)		(2.10)	
Posteriors				
VaccineKnowledge (V)		-6.47***		-1.87
		(1.98)		(1.39)
MisinformationSkills (M)		-0.27		2.12
		(2.18)		(1.49)
Control Mean	11.60	14.42	17.65	17.75
Control SD	23.62	26.78	17.02	18.58
N	1814	935	1814	935
Postestimation tests:				
p-value: $V_{Prior} = M_{Prior}$	0.06		<0.01	
p-value: $V_{Update} = M_{Update}$	0.42		0.34	
p-value: $V_{Posterior} = M_{Posterior}$		<0.01		<0.01

Each column presents treatment effect estimates relative to control with observations at the participant level. The dependent variables are indicated in the column title. In Columns 1 and 2, the dependent variable is the accuracy of participants' estimates regarding the percentage of young girls who would die from taking the HPV vaccine (correct answer: 0%). In Columns 3 and 4, the dependent variable is the accuracy of participants' estimates regarding the percentage of cervical cancers the HPV vaccine could prevent (correct answer: 90%). Accuracy is defined as the absolute difference between the correct answer and a participant's estimate, winsorized at the 95% level. Half of the participants responded to these questions after viewing two true and two false headlines on the HPV vaccine, while the other half did not view any headlines. In each column pair, the odd-numbered column includes all participants, and the even-numbered column includes only those who saw the headlines. Robust standard errors are in parentheses. All regressions control for community fixed-effects. Odd columns also control for whether the participant saw the headlines or not and include associated treatment interactions. Following a double-LASSO selection, we also control for whether the respondent completed some post-secondary education (regression 3), the respondent's reading level (regression 1), whether the respondent passed an initial attention check (regression 1), whether the respondent refused to share their monthly income (regression 1, 3).

* p<0.10, ** p<0.05, *** p<0.01

Table 4: Sharing Outcomes

	(1)	(2)	(3)	(4)	(5)
		Index Components			
	Sharing Discernment	Percentage of True Headlines Shared	Percentage of False Headlines Shared	Shared Any Headline	Number of Headlines Shared
VaccineKnowledge	0.05	0.01	-0.04	-0.01	-0.06
(V)	(0.03)	(0.02)	(0.03)	(0.01)	(0.07)
MisinformationSkills	0.06**	-0.01	-0.07**	-0.00	-0.15**
(M)	(0.03)	(0.02)	(0.03)	(0.01)	(0.07)
Control Mean	0.19	0.68	0.48	0.95	2.32
Control SD	0.57	0.35	0.50	0.21	1.29
N	1814	1814	1814	1814	1814
Postestimation tests:					
p-value: V=M	0.71	0.47	0.27	0.34	0.22

Each column presents treatment effect estimates relative to control with observations at the participant level. The dependent variables are indicated in the column title. For Column 1, the dependent variable is sharing discernment, measured as the percentage of true headlines shared (in Column 2) minus the percentage of false headlines shared (in Column 3). For Column 4, the dependent variable is a binary indicating whether the participant shared any headline. For Column 5, the dependent variable is the total number of headlines shared out of the four available. Robust standard errors are in parentheses. All regressions control for community and headline-set fixed-effects. Additional controls selected via double-LASSO selection are as follows: the respondent's reading level (Column 1), whether the respondent refused to share their monthly income (Columns 1, 2, 3, 4, 5).

* p<0.10, ** p<0.05, *** p<0.01

Table 5: Health Outcomes, Endline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Healthcare Utilization Index Components†				
	Vaccinated	Healthcare Utilization Index (HUI)	No. of visits to Doctor, Nurse or Clinic (pos. coded)	No. of purchases from a Pharmacy (pos. coded)	No. of purchases from a Marketer (neg. coded)	No. of purchases from a Peddler (neg. coded)	No. of purchases from/seen on Internet (neg. coded)
VaccineKnowledge	0.02	0.02	-0.06	-0.05	-0.09***	0.00	-0.00
(V)	(0.03)	(0.06)	(0.06)	(0.08)	(0.03)	(0.03)	(0.01)
	[1.00]	[1.00]					
MisinformationSkills	0.04	-0.01	-0.09*	-0.11	-0.02	-0.02	-0.01
(M)	(0.03)	(0.06)	(0.05)	(0.08)	(0.03)	(0.03)	(0.01)
	[1.00]	[1.00]					
Control Mean	0.27	0.00	0.41	1.41	0.26	0.18	0.02
Control SD	0.45	1.00	0.95	1.34	0.61	0.53	0.15
N	1644	1644	1644	1644	1644	1644	1644
Postestimation tests:							
p-value: V=M	0.38	0.56	0.63	0.43	0.02	0.38	0.31
q-value: V=M	[1.00]	[1.00]					

Each column presents treatment effect estimates relative to control with observations at the participant level. The dependent variables are indicated in the column title. In Column 1, the dependent variable is whether the participant got vaccinated at Endline. In Column 2, the dependent variable is the Healthcare Utilization Index (HUI).The HUI is the inverse-covariance weighted index of whether, and how many times, the participant visited a doctor/nurse/clinic or made purchases from a pharmacy (positively coded), and whether, and how many times, they made purchases from a marketer, peddler, internet (negatively coded). The continuous variables were winsorized at the 99%-level. In Columns 3 to 7, we look at the number of visits/purchases for each of the aforementioned categories individually (winsorized at the 99% level). Robust standard errors are in parentheses and FDR-adjusted p-values following the two-stage approach in Benjamini, Krieger, and Yekutieli (2006) are in square brackets. All regressions control for community fixed-effects as well as for the baseline value of the dependent variable. In Column 1, we also control for whether the participant was asked further question about its vaccine belief. Additional controls selected via double-LASSO selection (excluding the components of the HUI) are as follows: age (Column 4), whether the respondent's baseline/midline survey was lost due to an SCTO error (Columns 1, 5, 6), whether the respondent refused to share their monthly income (Columns 1, 2, 3, 4, 5, 6, 7), whether the respondent did not answer to the Baseline attention check (Column 5).

* p<0.10, ** p<0.05, *** p<0.01

† For concision, we only show the continuous subscomponents of the index. All the binary variables indicating whether the participant made at least one purchase/one visit are omitted.

References

Ali, A. and I. A. Qazi (2023). Countering misinformation on social media through educational interventions: Evidence from a randomized experiment in pakistan. *Journal of Development Economics 163*, 103108.

Amar, P., S. Badrinathan, S. Chauchard, and F. Sichart (2025). Countering misinformation early: Evidence from a classroom-based field experiment in india. Working paper.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484), 1481–1495.

Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in india. *American Political Science Review 115*(4), 1325–1341.

Bau, N., D. J. Henning, C. Low, and B. Steinberg (2024). Family planning, now and later: Infertility fears and contraceptive take-up. Technical report, CEPR Discussion Papers.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika 93*(3), 491–507.

Cepeda, N. J., H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin 132*(3), 354.

Chan, M.-p. S., C. R. Jones, K. Hall Jamieson, and D. Albarracín (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological science 28*(11), 1531–1546.

Danz, D., L. Vesterlund, and A. J. Wilson (2022). Belief elicitation and behavioral incentive compatibility. *American Economic Review 112*(9), 2851–2883.

Dupas, P. (2011). Do teenagers respond to hiv risk information? evidence from a field experiment in kenya. *American Economic Journal: Applied Economics 3*(1), 1–34.

Fazio, L., D. Rand, S. Lewandowsky, M. Susmann, A. J. Berinsky, A. Guess, P. Kendeou, B. Lyons, J. Miller, E. Newman, et al. (2024). Combating misinformation: a megastudy of nine interventions designed to reduce the sharing of and belief in false and misleading headlines.

Garimella, K. and S. Chauchard (2024). Whatsapp explorer: A data donation tool to facilitate research on whatsapp. *arXiv preprint arXiv:2404.01328*.

Guess, A. M., M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar (2020). A digital media literacy intervention increases discernment between mainstream and false news in the united states and india. *Proceedings of the National Academy of Sciences 117*(27), 15536–15545.

Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya (2023, October 29). Curtailing false news, amplifying truth. *SSRN Electronic Journal*.

Hagg, E., V. S. Dahinten, and L. M. Currie (2018). The emerging use of social media for health-related purposes in low and middle-income countries: A scoping review. *International journal of medical informatics 115*, 92–105.

Hoes, E., B. Aitken, J. Zhang, T. Gackowski, and M. Wojcieszak (2024). Prominent misinformation interventions reduce misperceptions but increase scepticism. *Nature Human*

Behaviour, 1–9.

Jalan, J. and E. Somanathan (2008). The importance of being informed: Experimental evidence on demand for environmental quality. *Journal of Development Economics 87*(1), 14–28.

Kozyreva, A., P. Lorenz-Spreen, S. M. Herzog, U. K. Ecker, S. Lewandowsky, R. Hertwig, A. Ali, J. Bak-Coleman, S. Barzilai, M. Basol, et al. (2024). Toolbox of individual-level interventions against online misinformation. *Nature Human Behaviour*, 1–9.

Larson, H. J. (2020). *Stuck: How Vaccine Rumors Start–and Why They Don't Go Away*. Oxford University Press.

Larson Williams, A., R. Mitrovich, L. Mwananyanda, and C. Gill (2019). Maternal vaccine knowledge in low-and middle-income countries—and why it matters. *Human vaccines & immunotherapeutics 15*(2), 283–286.

Lewandowsky, S. and S. Van Der Linden (2021). Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology 32*(2), 348–384.

Ludwig, J., S. Mullainathan, and A. Rambachan (2025). Large language models: An applied econometric framework. Technical report, National Bureau of Economic Research.

Lyons, B. A., J. M. Montgomery, A. M. Guess, B. Nyhan, and J. Reifler (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences 118*(23), e2019527118.

Madajewicz, M., A. Pfaff, A. Van Geen, J. Graziano, I. Hussein, H. Momotaj, R. Sylvi, and H. Ahsan (2007). Can information alone change behavior? response to arsenic contamination of groundwater in bangladesh. *Journal of Development Economics 84*(2), 731–754.

Meriggi, N. F., M. Voors, M. Levine, V. Ramakrishna, D. M. Kangbai, M. Rozelle, E. Tyler, S. Kallon, J. Nabieu, S. Cundy, et al. (2024). Last-mile delivery increases vaccine uptake in sierra leone. *Nature 627*(8004), 612–619.

Nan, X., Y. Wang, and K. Thier (2022). Why do people believe health misinformation and who is at risk? a systematic review of individual differences in susceptibility to health misinformation. *Social Science & Medicine 314*, 115398.

Nyhan, B., E. Porter, J. Reifler, and T. J. Wood (2020). Taking fact-checks literally but not seriously? the effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political behavior 42*, 939–960.

Pennycook, G., Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand (2021). Shifting attention to accuracy can reduce misinformation online. *Nature 592*(7855), 590–595.

Pennycook, G. and D. G. Rand (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature communications 13*(1), 2333.

Pliny (77). *Naturalis Historia*, Volume multiple volumes, varies by edition of *Loeb Classical Library*. Cambridge, MA: Harvard University Press.

Roediger, H. L. and A. C. Butler (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences 15*(1), 20–27.

Roozenbeek, J., S. Van Der Linden, B. Goldberg, S. Rathje, and S. Lewandowsky (2022). Psychological inoculation improves resilience against misinformation on social media. *Science advances 8*(34), eabo6254.

Solís Arce, J. S., S. S. Warren, N. F. Meriggi, A. Scacco, N. McMurry, M. Voors, G. Syunyaev,

A. A. Malik, S. Aboutajdine, O. Adeojo, et al. (2021). Covid-19 vaccine acceptance and hesitancy in low-and middle-income countries. *Nature medicine 27*(8), 1385–1394.

Taavila, S. H., M. Amara, K. Draginić, N. S. Jung, X. Li, and P. R. Sloksnath. Knowing what we don't know: Targeting overconfidence to combat misinformation. *Available at SSRN 4872927*.

Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in news. *American Economic Journal: Microeconomics 16*(2), 1–38.

Online Appendix A. Supplemental Tables and Figures

Table A1: Recruitment Funnel

	(1) Number of Participants	(2) Relative Share (/previous row)	(3) Total Share (/approached)
Successfully Approached	9367	100.0%	100.0%
Correct Enumerator Approach	9272	99.0%	99.0%
Consented to Recruitment Interview	7739	83.5%	82.6%
18 years old or older	7115	91.9%	76.0%
Lives in Western Urban Area	6732	94.6%	71.9%
Can read	5810	86.3%	62.0%
Has a phone	5265	90.6%	56.2%
Has used WA in the last 5 days	4898	93.0%	52.3%
Eligible for a dose of COVID-19 vaccine	4502	91.9%	48.1%
Has not already been interviewed	4494	99.8%	48.0%
Does not know anyone doing the survey (Eligible for B/M)	4482	99.7%	47.8%
Interviewed	1879	41.9%	20.1%
Eligible for analysis (Baseline and/or Midline)	1823	97.0%	19.5%

The table presents the recruitment funnel from approached to completed baseline/midline. Fifty six participants interviewed by one enumerator who was let go from the study mid data collection due to quality concerns are not included in the analytical sample.

Table A2: Baseline Summary Statistics and Balance

	(1)	(2)	(3)	(4)	(5)
	Mean (SE)				
	Full Sample	Control	Vaccine	Misinformation	p-value for F-test of joint significance of treatments, after regression of row variable on treatment
Panel A: Demographics					
Woman	0.25	0.22	0.26	0.26	0.17
	(0.01)	(0.02)	(0.02)	(0.02)	
Age	24.61	24.99	24.26	24.58	0.25
	(0.15)	(0.28)	(0.26)	(0.26)	
Income below Median	0.52	0.50	0.57	0.50	0.01
	(0.01)	(0.02)	(0.02)	(0.02)	
Reading Level Index (/100)	73.34	73.62	72.44	73.93	0.39
	(0.51)	(0.88)	(0.90)	(0.86)	
Reached post-secondary education	0.43	0.43	0.41	0.45	0.37
	(0.01)	(0.02)	(0.02)	(0.02)	
Panel B: Health					
Healthcare Utilization Index	0.03	0.00	0.03	0.05	0.81
	(0.02)	(0.04)	(0.04)	(0.04)	
Took the COVID vaccine	0.11	0.13	0.10	0.10	0.13
	(0.01)	(0.01)	(0.01)	(0.01)	
Panel C: Digital Literacy and Engagement					
Digital Literacy Score (/3)	2.34	2.34	2.33	2.36	0.83
	(0.02)	(0.03)	(0.03)	(0.03)	
Uses WA *many* times a day	0.85	0.85	0.83	0.86	0.48
	(0.01)	(0.01)	(0.02)	(0.01)	
Nr of WA conversation in the last two days	15.50	15.57	15.49	15.45	0.90
	(0.34)	(0.60)	(0.65)	(0.51)	
N	1823	604	595	624	

Columns 1 to 4 present baseline means with standard errors in parentheses for the full sample (Column 1) and each study arm (Columns 2-4). The variables in row 1 and 2 are the participant's gender and age, respectively. The variable in row 3 indicates whether the participant's monthly income is below the median of our sample: 500 Leones per month. Participants reported their income by selecting one of the following categories: 250 Leones or less, between 250 and 500 Leones, between 500 and 1000 Leones, between 1000 and 1500 Leones, or more than 1500 Leones. The variable in row 4, the reading level index, corresponds to the average of the reading score given by the enumerators at Recruitment & Baseline, rescaled from a 4-point scale (beginner, basic, intermediate, proficient) to 0-100). The variable in row 5, reached post-secondary education, corresponds to whether the participant declared having completed or being currently studying at the Technical/Vocational/College or University or Master's level. The variable in row 6, the Healthcare Utilization Index, is the inverse-covariance weighted index of whether, and how many times, the participant visited a doctor/nurse/clinic or made purchases from a pharmacy (positively coded), and whether, and how many times, they made purchases from a marketer, peddler, internet (negatively coded), with all continuous variables winsorized at the 99%-level. The variable in row 7, took the COVID vaccine, indicates whether the participant took any COVID vaccine before baseline. The variable in row 8, digital literacy score, indicates how many given WhatsApp features the participant declared knowing how to use out of three. The variable in row 9, uses WhatsApp many times a day, indicating whether the participant declared using WhatsApp 'Many times a day' (as opposed to 'A few times a day', 'Once a day', 'Once a week' or 'Every month or less'). The variable in row 10, number of WhatsApp conversations in the last two days indicates how many people & groups the participant declared having communicated with on WhatsApp in the last two days (winsorized at the 99%-level). Each cell in Column 5 provides the p-value from an F-test on the joint significance of the two treatment variables, from an OLS regression of the row variable on the treatment assignment dummies with community fixed-effects.

Table A3: Attrition

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean (SD)					
	Full Sample	Control	Vaccine	Misinformation	p-value for F-test of joint significance of treatments, after regression of row variable on treatment	p-value of F-test of joint significance of interaction terms, after regression of completion on treatment, row variables and interactions
Panel A: Endline Completion Rates						
Interview Completed	0.90	0.90	0.89	0.91	0.49	
	(0.01)	(0.01)	(0.01)	(0.01)		
Interview Completed in the Lab	0.60	0.60	0.59	0.59	0.92	
	(0.01)	(0.02)	(0.02)	(0.02)		
Panel B: Baseline Statistics for Endline Sample						
General Demographics						
Woman	0.24	0.20	0.26	0.26	0.05	
	(0.01)	(0.02)	(0.02)	(0.02)		
Age	24.60	24.87	24.35	24.57	0.63	
	(0.16)	(0.30)	(0.28)	(0.27)		
Income below Median	0.52	0.51	0.58	0.49	0.01	
	(0.01)	(0.02)	(0.02)	(0.02)		
Reading Level Index (/100)	73.42	74.13	72.49	73.61	0.39	
	(0.53)	(0.91)	(0.94)	(0.90)		
Reached post-secondary education	0.43	0.43	0.41	0.45	0.42	
	(0.01)	(0.02)	(0.02)	(0.02)		
Health						
Healthcare Utilization Index	0.03	0.02	0.02	0.05	0.92	
	(0.02)	(0.04)	(0.05)	(0.04)		
Took the COVID vaccine	0.11	0.14	0.10	0.09	0.04	
	(0.01)	(0.01)	(0.01)	(0.01)		
Digital Literacy and Engagement						
Digital Literacy Score (/3)	2.35	2.34	2.33	2.37	0.69	
	(0.02)	(0.03)	(0.03)	(0.03)		
Uses WA *many* times a day	0.85	0.85	0.83	0.86	0.51	
	(0.01)	(0.02)	(0.02)	(0.01)		
Nr of WA conversation in the last two days	15.80	15.89	15.77	15.74	0.89	
	(0.36)	(0.64)	(0.71)	(0.54)		
p-value: treatments X all variables above						0.25
N	1644	543	532	569	1644	1823

Columns 1 to 4 present baseline means with standard errors in parentheses for the full sample (Column 1) and each study arm (Columns 2-4). The second panel restricts the sample to participants who completed Endline. The variable in row 1 and 2 indicates whether the participant completed the endline survey and whether they did so in one of our labs (as opposed to on the field). The variables in row 3 and 4 are the participant's gender and age, respectively. The variable in row 5 indicates whether the participant's monthly income is below the median of our sample: 500 Leones per month. Participants reported their income by selecting one of the following categories: 250 Leones or less, between 250 and 500 Leones, between 500 and 1000 Leones, between 1000 and 1500 Leones, or more than 1500 Leones. The variable in row 6, the reading level index, corresponds to the average of the reading score given by the enumerators at Recruitment & Baseline, rescaled from a 4-point scale (beginner, basic, intermediate, proficient) to 0-100). The variable in row 7, reached post-secondary education, corresponds to whether the participant declared having completed or being currently studying at the Technical/Vocational/College or University or Master's level. The variable in row 8, the Healthcare Utilization Index, is the inverse-covariance weighted index of whether, and how many times, the participant visited a doctor/nurse/clinic or made purchases from a pharmacy (positively coded), and whether, and how many times, they made purchases from a marketer, peddler, internet (negatively coded), with all continuous variables winsorized at the 99%-level. The variable in row 9, took the COVID vaccine, indicates whether the participant took any COVID vaccine before baseline. The variable in row 10, digital literacy score, indicates how many given WhatsApp features the participant declared knowing how to use out of three. The variable in row 11, uses WhatsApp many times a day, indicates whether the participant declared using WhatsApp 'Many times a day' (as opposed to 'A few times a day', 'Once a day', 'Once a week', or 'Every month or less'). The variable in row 12, number of WhatsApp conversations in the last two days, indicates how many people & groups the participant declared having communicated with on WhatsApp in the last two days (winsorized at the 99%-level). Each cell in Column 5 provides the p-value from an F-test on the joint significance of the two treatment variables, from an OLS regression of the row variable on the treatment assignment dummies with community fixed-effects. The single value in Column 6 provides the p-value from an F-test of all interaction terms, from an OLS regression of a binary variable indicating whether the participant completed endline on treatment assignment dummies, row variables, and row variables interacted with treatment assignment dummies, controlling for community fixed-effects.

Table A4: Midline, Experimental & Real Headline Comparison

	(1)	(2)	(3)	(4)	(5)	(6)
	Real Headline Discernment for			Experimental Headlines Discernment for		
	All Headlines	True Headlines	False Headlines	True & False Negative Headlines	True Headlines	False Negative Headlines
VaccineKnowledge	0.08***	0.10***	0.06	0.14***	0.13***	0.15***
(V)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.05)
MisinformationSkills	0.08***	0.03	0.13***	0.12***	0.08**	0.21***
(M)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
Control Mean	0.00	0.00	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00	1.00	1.00
Unstandardized Control Mean	74.82	83.74	65.89	82.68	88.97	70.10
Unstandardized Control SD	35.85	30.08	38.81	30.68	24.85	36.77
N	14512	7256	7256	10890	7260	3630
Number of Participants	1814	1814	1814	1815	1815	1815
Postestimation tests:						
p-value: V=M	0.92	0.03	0.07	0.52	0.10	0.18

Each column presents treatment effect estimates relative to control with observations at the headline level. The dependent variables is the midline headline discernment index for different subset of headlines: all real headlines in Column 1, true real headlines in Column 2, false real headlines in Column 3 (which are de-facto negative), true and false negative experimental headlines in Column 4, true experimental headlines in Column 5, false negative experimental headlines in Column 6. This metric captures how accurately participants assessed the probability of a headline being entirely true. If the headline was actually true, their score corresponds directly to their assessment. Conversely, if the headline was false, their score is calculated as 100 minus their assessment. Standard errors clustered at the individual level are in parentheses. All regressions control for community and item fixed-effects as well as for module order. Additional controls selected via double-LASSO selection are as follows: the respondent's gender (Column 5), whether the respondent is engaged in post-secondary education (Column 6), whether the respondent knows how to report someone on WA (Columns 1, 3, 4, 6), the respondent's reading level (Columns 1, 3, 4, 6), whether the respondent passed an initial attention check (Columns 1, 3), whether the respondent refused to share their monthly income (Columns 1, 2, 3, 4, 5, 6).

* p<0.10, ** p<0.05, *** p<0.01

Table A5: Discernment, Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Discernment - Item Level:				Discernment - Participant Level:			
	Standardized	Raw	Accuracy ($1 \sim =100\%$) ($0 \sim \leq 75\%$)	Accuracy ($1 \sim \geq 75\%$) ($0 \sim \leq 50\%$)	Standardized	Raw	Accuracy ($1 \sim =100\%$) ($0 \sim \leq 75\%$)	Accuracy ($1 \sim \geq 75\%$) ($0 \sim \leq 50\%$)
Panel A: Experimental Headlines								
VaccineKnowledge (V)	0.09***	2.54***	0.08***	0.05***	0.16***	2.47***	0.08***	0.05***
	(0.03)	(0.87)	(0.01)	(0.01)	(0.06)	(0.89)	(0.01)	(0.01)
MisinformationSkills (M)	0.18***	6.20***	0.08***	0.07***	0.39***	6.21***	0.08***	0.07***
	(0.03)	(0.83)	(0.01)	(0.01)	(0.05)	(0.84)	(0.01)	(0.01)
Control Mean	0.00	77.75	0.65	0.70	0.00	77.75	0.65	0.70
Control SD	1.00	34.38	0.48	0.46	1.00	15.79	0.25	0.21
N	14520	14520	14520	14520	1815	1815	1815	1815
Number of Participants	1815	1815	1815	1815	1815	1815	1815	1815
Postestimation tests:								
p-value: V=M	<0.01	<0.01	0.55	0.05	<0.01	<0.01	0.56	0.05
Panel B: Real Headlines								
VaccineKnowledge (V)	0.08***	2.68***	0.06***	0.05***	0.16***	2.52***	0.06***	0.05***
	(0.02)	(0.84)	(0.01)	(0.01)	(0.06)	(0.85)	(0.01)	(0.01)
MisinformationSkills (M)	0.08***	2.29***	0.06***	0.05***	0.15***	2.32***	0.06***	0.05***
	(0.02)	(0.80)	(0.01)	(0.01)	(0.05)	(0.82)	(0.01)	(0.01)
Control Mean	0.00	74.82	0.61	0.66	0.00	74.82	0.61	0.66
Control SD	1.00	35.85	0.49	0.47	1.00	15.42	0.25	0.22
N	14512	14512	14512	14512	1814	1814	1814	1814
Number of Participants	1814	1814	1814	1814	1814	1814	1814	1814
Postestimation tests:								
p-value: V=M	0.63	0.92	0.62	0.76	0.82	0.82	0.78	0.60
Panel C: Statements								
VaccineKnowledge (V)	0.16***	5.84***	0.11***	0.09***	0.38***	5.81***	0.11***	0.09***
	(0.02)	(0.81)	(0.01)	(0.01)	(0.05)	(0.82)	(0.01)	(0.01)
MisinformationSkills (M)	0.03	1.49*	0.01	0.03**	0.10*	1.46*	0.01	0.02**
	(0.02)	(0.81)	(0.01)	(0.01)	(0.05)	(0.82)	(0.01)	(0.01)
Control Mean	0.00	62.34	0.48	0.53	0.00	62.34	0.48	0.53
Control SD	1.00	41.08	0.50	0.50	1.00	15.29	0.22	0.20
N	23584	23584	23584	23584	1815	1815	1815	1815
Number of Participants	1815	1815	1815	1815	1815	1815	1815	1815
Postestimation tests:								
p-value: V=M	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01

Each column presents treatment effect estimates relative to control, with observations at the headline (Columns 1 to 4) or participant (Columns 5 to 8) level. The dependent variables are indicated in the column headers: in Column 1, the standardized discernment index (as pre-registered); in Column 2, the unstandardized discernment index; in Column 3, the proportion of news items correctly identified, treating as correct any false news rated 0% likely to be true and true news rated 100% likely to be true; in Column 4, the proportion of news items correctly identified, treating as correct any false news rated 0% or 25% likely to be true and true news rated 75% or 100% likely to be true. Columns 5 to 8 present the same outcomes at the participant level. Panel A shows results for experimental headlines, Panel B for real-world headlines, and Panel C for statements. Standard errors clustered at the individual level (Columns 1–4) and robust standard errors (Columns 5–8) are in parentheses. All regressions include community fixed effects; Columns 1–4 also include headline fixed effects. Panel A regressions further control for module order. Following a double-LASSO procedure, we also control for the respondent's gender (Columns 1–8), whether the respondent completed some post-secondary education (Columns 3, 7), whether the respondent is engaged in post-secondary education (Column 1), whether the respondent knows how to report someone on WA (Columns 1–8 in all panels), the respondent's reading level (Columns 1–8 in all panels), whether the respondent passed an initial attention check (Columns noted for each panel), and whether the respondent refused to share income (Columns 1–8 in all panels).

* p<0.10, ** p<0.05, *** p<0.01

Table A6: Midline, Experimental Headline Discernment, by headline type

	(1) Core: Source only	(2) Core: Source & Additional Taught Feature	(3) Adjacent: Source & Untaught Features	(4) False with Fake Source	(5) True without Source
VaccineKnowledge	0.28***	0.00	0.03	0.02	0.13**
(V)	(0.09)	(0.05)	(0.05)	(0.06)	(0.05)
MisinformationSkills	0.32***	0.26***	0.27***	-0.05	-0.15**
(M)	(0.08)	(0.05)	(0.04)	(0.06)	(0.06)
Control Mean	0.00	0.00	0.00	0.00	0.00
Control SD	1.00	1.00	1.00	1.00	1.00
Unstandardized Control Mean	55.86	65.37	69.98	68.98	94.77
Unstandardized Control SD	39.57	39.09	37.63	37.19	20.08
N	839	2791	3630	1815	1815
Number of Participants	729	1705	1815	1815	1815
Postestimation tests:					
p-value: V=M	0.65	<0.01	<0.01	0.28	<0.01
p-value: $V_{Col\ 1} = V_{Col\ 2}$		<0.01			
p-value: $M_{Col\ 1} = M_{Col\ 2}$		0.57			
p-value: $V_{Col\ 2} = V_{Col\ 3}$			0.74		
p-value: $M_{Col\ 2} = M_{Col\ 3}$			0.86		

Each column presents treatment effect estimates relative to the control group, using observations at the headline level. The dependent variable is the midline experimental headline discernment index, calculated for various subsets of headlines. These include: false core headlines characterized solely by the absence of a source (Column 1); false core headlines featuring the absence of a source and the inclusion of emotional language, a fake doctor, or a mix of true and false information (Column 2); false adjacent headlines with the absence of a source and the inclusion of untaught features (Column 3); false headlines with a fake source (Column 4); and true headlines without a source (Column 5). Standard errors clustered at the individual level are in parentheses. All regressions control for community and item fixed-effects as well as for module order. Additional controls selected via double-LASSO selection are as follows: whether the respondent completed some post-secondary education (Column 3), whether the respondent is engaged in post-secondary education (Columns 2, 3), whether the respondent knows how to report someone on WA (Columns 2, 3, 4), the respondent's reading level (Columns 2, 3, 4), whether the respondent passed an initial attention check (Column 3), whether the respondent refused to share their monthly income (Columns 3, 4, 5).

* p<0.10, ** p<0.05, *** p<0.01

Table A7: Endline Outcomes, IPW

	(1) Statement Discernment	(2) Experiment Headline Discernment	(3) Real Headline Discernment	(4) Vaccinated	(5) Healthcare Utilization Index
Vaccine	0.10***	0.21***	0.13***	0.02	0.03
Treatment	(0.02)	(0.03)	(0.03)	(0.03)	(0.06)
Misinformation	0.03	0.22***	0.20***	0.04	-0.01
Treatment	(0.02)	(0.03)	(0.03)	(0.03)	(0.06)
Control Mean	0.00	0.00	0.00	0.27	0.00
Control SD	1.00	1.00	1.00	0.45	1.00
Unstandardized Control Mean	59.00	67.86	65.89		
Unstandardized Control SD	40.35	38.30	37.18		
N	21369	6571	6570	1644	1644
Number of Participants	1644	1643	1643	1644	1644
Postestimation tests:					
p-value: V=M	<0.01	0.75	0.02	0.36	0.50

Each column presents inverse probability weight (IPW) treatment effect estimates relative to control with observations at the statement (Column 1), headline (Columns 2 and 3) and participant (Columns 4 and 5) level. To calculate the weights, we predicted endline completion usinf the following baseline variables: Gender, Age, Income Below Median, Reading Level, Post-Secondary Education, Healthcare Utilization Index, Took Vaccine at Baseline, Uses WhatsApp Many Times a Day, Number of WhatsApp Conversations, Occupation and Baseline community. The dependent variables are indicated in the column title. For column 1, the dependent variable is the statement index, measured on a 0 to 100 scale. This metric captures how accurately participants assessed the probability of a statement being true. If the statement was actually true, their score corresponds directly to their assessment. Conversely, if the statement was false, their score is calculated as 100 minus their assessment. For Columns 2 and 3, the dependent variable is the headline discernment for experimental and real headlines respectively, measured on a 0 to 100 scale. Scores were calculated in the same manner as described for the statement discernment metric. For Column 4, the dependent variable is whether the participant got vaccinated at Endline. For Column 5, the dependent variable is the Healthcare Utilization Index (HUI).The HUI is the inverse-covariance weighted index of whether, and how many times, the participant visited a doctor/nurse/clinic or made purchases from a pharmacy (positively coded), and whether, and how many times, they made purchases from a marketer, peddler, internet (negatively coded). The continuous variables were winsorized at the 99%-level. Standard errors clustered at the individual level (Columns 1 to 3) and robust standard errors (Columns 4 and 5) are in parentheses. All regressions control for community fixed-effects. Columns 1, 2, and 3 also control for item fixed-effects and module order, Column 4 for the baseline HUI, and column 5 for the baseline vaccine status of the participant and for whether the participant was asked further question about its vaccine belief. Additional controls selected via double-LASSO selection are as follows: whether the respondent is engaged in post-secondary education (Columns 1, 2, 3), whether the respondent knows how to report someone on WA (Columns 1, 2), whether the respondent uses google at least once a day (Column 3), the respondent's reading level (Columns 1, 2, 3), whether the respondent passed an initial attention check (Column 1), whether the respondent's baseline/midline survey was lost due to an SCTO error (Columns 2, 3, 4), whether the respondent refused to share their monthly income (Columns 1, 2, 3, 4, 5).

* p<0.10, ** p<0.05, *** p<0.01

Table A8: Outcomes from Family 3

	(1) Real Headline Discernment	(2) Sharing Discernment Index	(3) Posterior Learning Accuracy
VaccineKnowledge (V)	0.08***	0.04	-3.12***
	(0.02)	(0.03)	(0.97)
	[<0.01]	[0.17]	[<0.01]
MisinformationSkills (M)	0.08***	0.06**	0.64
	(0.02)	(0.03)	(1.03)
	[<0.01]	[0.04]	[0.36]
Control Mean	0.00	0.19	13.76
Control SD	1.00	0.57	12.93
Unstandardized Control Mean	74.82		
Unstandardized Control SD	35.85		
N	14512	1814	935
Number of Participants	1814	1814	935
Postestimation tests:			
p-value: V=M	0.92	0.47	<0.01
q-value: V=M	0.50	0.36	<0.01

Each column presents treatment effect estimates relative to control with observations at the headline (Column 1) or participant (Columns 2 and 3) level. The dependent variables are indicated in the column title. For Column 1, the dependent variable is the real headline discernment, measured on a 0 to 100 scale. This metric captures how accurately participants assessed the probability of a real headline being entirely true. If the headline was actually true, their score corresponds directly to their assessment. Conversely, if the headline was false, their score is calculated as 100 minus their assessment. For Column 2, the dependent variable is the sharing discernment, measured as the percentage of true headlines shared minus the percentage of false headlines shared. For Column 1, the dependent variable is the learning accuracy. The participants had to answer the following two questions: 'What percentage of young girls taking the HPV vaccine will die because they took the vaccine?' & 'What percentage of cervical cancers could be prevented by the HPV vaccine?' (correct answers: 0% & 90%, respectively). Half of the participants were asked these two questions after seeing some true and false headlines on the HPV vaccine (2 true & 2 false) while half were not shown any headlines. In Column 3, we conduct the regression only for participants who did see the headlines. The index is the average of the accuracy, the difference between the correct answer and the participant's estimate, received by the participant on the two questions (winsorized at 95%). Standard errors clustered at the individual level (Column 1) and robust standard errors (Columns 2 to 3) are in parentheses; FDR-adjusted p-values following the two-stage approach in Benjamini, Krieger, and Yekutieli (2006) are in square brackets. All regressions control for community fixed-effects. Column 1 also control for headline fixed-effects, Column 2 for headline-set fixed-effects. Additional controls selected via double-LASSO selection are as follows: whether the respondent knows how to report someone on WA (Columns 1, 2), the respondent's reading level (Column 1), whether the respondent passed an initial attention check (Column 1), whether the respondent refused to share their monthly income (Columns 1, 2).

* p<0.10, ** p<0.05, *** p<0.01

Online Appendix B. Intervention and Survey Materials

B.1 Script for Vaccine Knowledge Training Video

People get colds and illnesses all the time. And if they're mild, our body has the defenses to fight them off. But there are some more dangerous diseases that we can't fight off on our own. So we need to prepare for them before we become infected. And that's where vaccines come in.

Vaccines don't cure us if we already have the disease. But they help protect us against diseases we may encounter in the future. And here's how they do it: They show our bodies what the dangerous disease looks like and issue a warning. Our bodies can then recognize the disease and create an army strong enough to defeat it. If and when we become infected with this disease, our body's soldiers are then ready to go into battle with the right weapons. Creating this new army can be tiring for the body, so sometimes people can feel a little unwell for a day or two after receiving the vaccine. But this soon fades and our body stays strong and protected.

There are many different vaccines to protect us against different illnesses and we can get vaccinated at any age. It is usually necessary to have two or more doses to become fully protected.

The vaccines used in Sierra Leone have been approved by the WHO and the Ministry of Health and Sanitation. This means they have been thoroughly tested before we receive them to make sure they work and that they're safe.

Getting vaccinated is the best way of ensuring that we and the people we love stay strong and healthy.

B.2 Script for Misinformation Skills Training Video

Every day we are bombarded with information and a lot of it just isn't true. But it's often really difficult to distinguish the true information from the false. Here are some clues that will help you to identify false information.

Messages designed to make us feel powerful emotions that cloud our judgment like fear. Photos or footage of people who look like experts but don't actually have any training or qualifications in the subject they claim to be an expert in. All that glitters is not gold. Advice that doesn't come from an organization you know and trust. And statements that mix some true information in with the false.

If we act on health information that isn't true, it can stop us from getting the treatment we need Or even make us do things that damage our health.

So when you come across something you're not sure about, ask yourself these questions:

Am I being manipulated into feeling alarmed or too happy? Have I ever heard of this so-called expert? Are they for real? Does this information come from an organization I trust like the Ministry of Health and Sanitation, the WHO or the SLBC or someone who hasn't checked their facts? And even if part of this sounds right, does the rest sound like fake news?

After asking yourself these questions, if you're still not sure whether something is true or not true, check with somebody who is knowledgeable, like a health worker. Misinformation is bad for you! Check your facts and only act on information you know you can trust.

B.3 Script for Control Training Video

Ibrahim: Guys, I need your help to build a tippy tap like the one you have at home. I want to surprise my mum.

John: Great idea Ibrahim!

Girl: Okay. John, you go and find two forked sticks and two straight sticks, and Ibrahim you go and get a container of clean water and soap. I'm going to collect some small pebbles for the pit.

Ibrahim: But where are we going to find a rope?

John: No, no, don't even think about using the cord from my shorts! They'll fall down, I don't want to end up naked!

Girl: That's what we need, the clothesline! We have to plant the two forked sticks in the ground, lay a straight stick across them, attach the container, link the pedal to the water container, then make a pile of pebbles and attach the soap. There! Wow, it's so easy to build!

Ibrahim: So cool!

Ibrahim's mum: Who took my clothesline? Show me what you're hiding!

Ibrahim: Mum, our teacher told us a tippy tap is very good for washing hands with soap because you don't have to touch the water container with dirty hands and it saves water too.

Ibrahim's mum: Wow, only 8 years old and you did this all by yourselves!

Girl: Auntie, I don't want Ibrahim to keep getting diarrhea all the time. I don't want him to end up in heaven like our older brother.

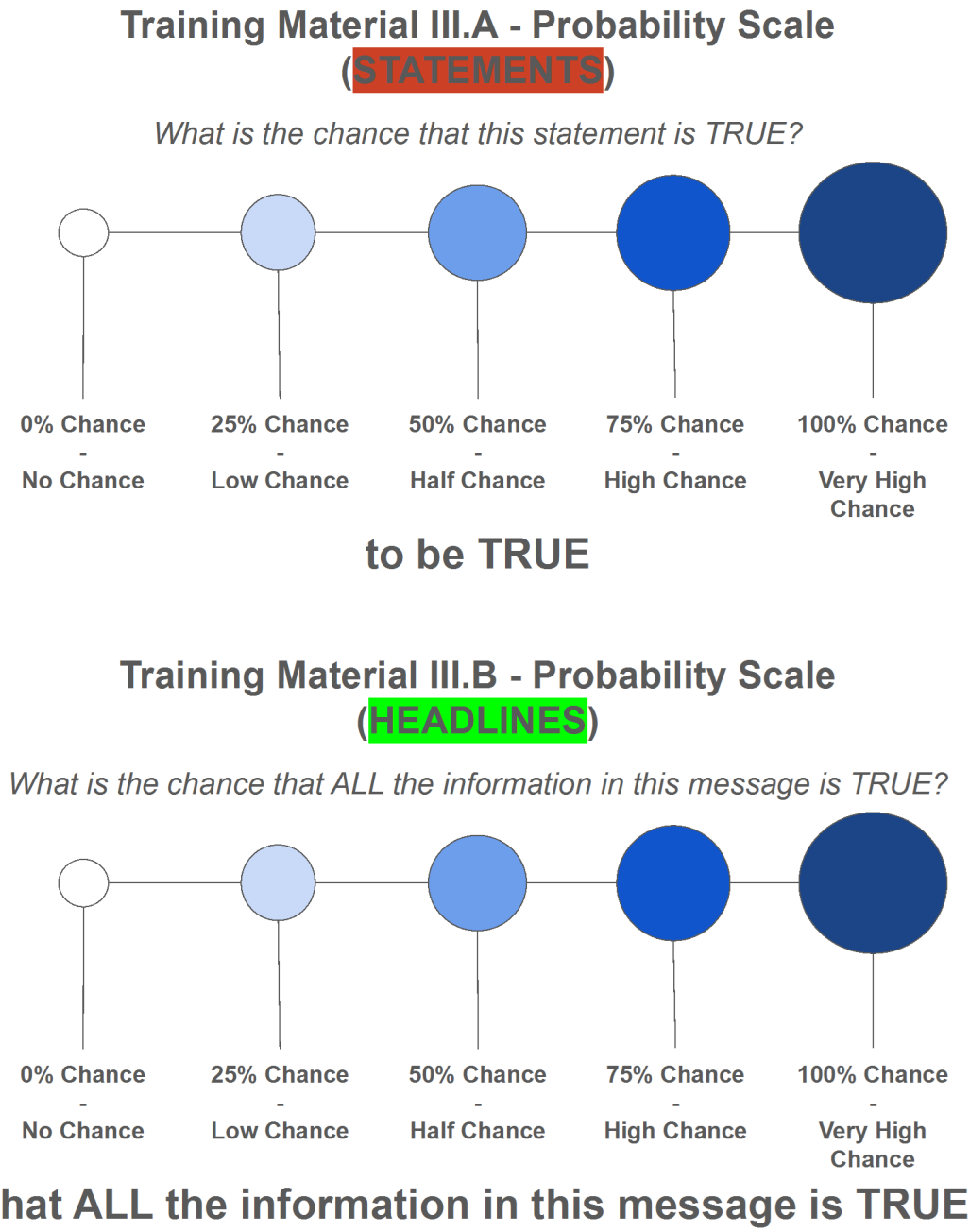
John: We can ask our mum to give you another rope...

Ibrahim's mum: Oh no, thank you John, my family's health is more important than a clothesline.

Girl and John: Well done Ibrahim!

The tippy tap is the best way to wash your hands safely and it's simple to build. Make one for your family!

Figure B1: Translated Headline & Statement Scale



Notes: Participants were shown the relevant scale in Krio when they had to evaluate whether statements or headlines were likely to be false.