

Pre-print version 5/12/26. This paper has not been peer-reviewed

Statistical Reasoning regarding Possible Adverse Outcomes of Vaccination

Alexander C. Walker^{1,*}, Jeremy D. Gretton^{1,*}, Ethan A. Meyers¹, Alexandra M. van der Valk^{1,2},

Jonathan A. Fugelsang¹, and Derek J. Koehler¹

¹ Department of Psychology, University of Waterloo, Waterloo, ON, Canada

² Geriatric Medicine Research, Nova Scotia Health, Halifax, NS, Canada

* Denotes Equal Contribution

Author Note

Alexander C. Walker  <https://orcid.org/0000-0003-1431-6770>

Jeremy D. Gretton  <https://orcid.org/0000-0002-2417-4035>

Ethan A. Meyers  <https://orcid.org/0000-0001-6171-6780>

Alexandra M. van der Valk  <https://orcid.org/0000-0001-5176-3516>

Jonathan A. Fugelsang  <https://orcid.org/0000-0002-6342-7023>

Derek J. Koehler  <https://orcid.org/0000-0002-3596-7882>

This research was supported by grants from The Natural Sciences and Engineering Research Council of Canada (to JAF and DJK). Correspondence concerning this article should be addressed to Alexander C. Walker, 200 University Avenue West, Waterloo, ON, Canada, N2L 3G1. Email: a24walke@uwaterloo.ca

Abstract

Reports of adverse events following vaccination can heighten concerns about vaccine safety. Across two experiments ($N = 602$ U.S residents), we examine how individuals evaluate evidence for a causal link between a vaccine and an adverse outcome, focusing on two key statistical considerations: the background incidence (or base rate) of the outcome in the absence of vaccination and the number of adverse outcomes assessed. Participants repeatedly judged the strength of evidence that a vaccine had caused an adverse outcome as information about a vaccine trial accumulated. Providing base rate information reduced perceived evidence strength and concerns about vaccine safety, particularly when the frequency of adverse outcomes aligned with population expectations, with these effects being more pronounced among more numerate individuals. Explicitly translating base rate information into expected case counts further reduced perceived risk, including among less numerate participants. Conversely, explaining how monitoring many outcomes increases the likelihood of false positives had little impact on participants' judgments, regardless of their level of numeracy. Together, these findings show how individuals' numerical abilities interact with the presentation of statistical information to shape vaccine-related judgments, highlighting both effective strategies for improving risk assessment and persistent challenges in communicating complex statistical principles.

Keywords: health communication, base rates, numeracy, vaccination, clustering illusion

Public Significance Statement

Reports of adverse health events following vaccination can increase public concern, in part because people often struggle to judge whether such events are caused by vaccination. This research shows that providing simple statistical information—such as how often an outcome occurs in the general population and how many cases would be expected in clinical trials by chance alone—can reduce unwarranted concern, even among individuals with lower numerical ability. However, it also shows that people, regardless of numerical ability, struggle to incorporate some statistical principles into their judgments, hindering risk assessment and posing an important challenge for public health communication.

Introduction

“It is important to stress that a health problem that arises after vaccination is not necessarily caused by vaccination. Millions of people will be vaccinated in the next nine months. After vaccination, some will have a heart attack or stroke, some will win the lottery, others will fall in love. And none of these events will be attributable to the vaccine.” – Guichon & Franco (2021)

On September 8th, 2020, AstraZeneca temporarily halted clinical trials of its COVID-19 vaccine following a possible serious adverse reaction to the vaccine by a trial participant, who had been diagnosed with transverse myelitis (inflammation of the spinal cord). At a time when the public was anticipating that various COVID-19 vaccines would soon be widely available, news that AstraZeneca had halted a clinical trial received considerable attention (Miller & Edwards, 2020; Walsh, 2020; Wu, 2020). Several days later, AstraZeneca completed its investigations and concluded that it was safe to resume clinical trials.

Vaccines save lives, yet vaccine hesitancy continues to pose a significant public health challenge (Jia et al., 2023; Lo & Hotez, 2017; Salmon et al., 2015). Concerns about vaccine safety, including the possibility of side effects or adverse reactions to vaccination, are a central driver of such hesitancy (Fisher et al., 2020; Neumann-Böhme et al., 2020; Smith et al., 2017; Solís Arce et al., 2021). Thus, it is critical to understand how members of the public respond to reports of adverse events following vaccination, such as that reported in the AstraZeneca clinical trial.

In reporting the initial case that led to the temporary halt of the clinical trial, *The New York Times* quoted a vaccine expert who warned against overreaction, noting that “The larger your study group, the more likely you’ll find an adverse event. This could occur spontaneously.” (Wu, 2020). This warning is, in effect, a prescription to use statistical reasoning when evaluating

possible adverse outcomes of vaccination. However, individuals vary widely in their statistical reasoning abilities, and even experts are prone to consequential errors in statistical judgment (Casscells et al., 1978; Gigerenzer et al., 2007; Hoffrage & Gigerenzer, 1998). The present research examines how individuals evaluate evidence for a potential causal link between a vaccine and an adverse outcome. To do this, we use a vignette in which cases of myelitis were observed in a set of clinical trial participants following vaccination.¹ The vignette was partitioned into multiple sequential steps (see Table 1), with each step introducing new information. After each step, participants evaluated both the strength of the evidence for a causal link between the vaccine and myelitis and their personal level of concern. This parceling reduced how much information had to be absorbed at each step and allowed us to isolate the impact of each piece of information on participants' judgments.

¹ The vaccine was described in general terms to avoid linking it to COVID-19 vaccinations, which were being made available to the public at the time this study was conducted (March through May of 2021).

Table 1

Experiment 1 Scenario

Step	(Additional) Information
Intro	There’s always concern that vaccines that are rolled out could have undesirable or adverse outcomes. So, during clinical trials and the eventual rollout of the vaccine to the wider public, public health authorities monitor for the occurrence of adverse outcomes. Their goal is to try to determine whether any observed adverse outcomes are actually caused by the vaccine.
1	Consider a scenario in which, during the initial public rollout of a new vaccine, there are confirmed reports of people developing myelitis after vaccination. Myelitis is a rare disease that involves inflammation and can lead to physical disability.
2	More specifically, suppose that [10 or 25] people developed myelitis after vaccination.
3	More specifically, in this scenario, suppose that 1,000 individuals in total had been vaccinated, out of which the [10 or 25] people mentioned previously developed myelitis.
4	Suppose that in the general population, even without vaccination, myelitis occurs in 1% of the population over any given 2-month period.
5	So, out of the 1,000 people who were vaccinated in the scenario, we would expect to observe 10 people develop myelitis, by chance alone, even if the vaccine does not cause myelitis.
6	Suppose that myelitis is just one of hundreds of potential rare adverse effects or illnesses (e.g., rashes, arthritis, stroke, encephalitis, asthma, tinnitus) that public health authorities monitored for among the 1,000 vaccinated individuals. Out of the many rare diseases that public health authorities watched for, suppose myelitis was one that occurred frequently enough to be ‘flagged’ (brought to the attention of those monitoring adverse effects or illnesses) among the 1,000 vaccinated individuals.

Note. At each step, pertinent information from previous steps was reiterated before new information was presented. This partitioning ensured that participants did not have to rely on memory for previously presented information. Square brackets indicate values that differed across randomly assigned Illness Frequency conditions in Experiment 1.

Base Rate Information and Evaluations of Vaccine Safety

Participants were first informed that some individuals in a clinical trial developed myelitis following vaccination (Step 1). Next, they were told how many cases had been observed

(10 or 25 in Experiment 1, depending on Illness Frequency condition; Step 2) and the total number of trial participants (1000; Step 3). On the basis of this information alone, it is difficult to assess whether the vaccine increased the risk of myelitis, as such an evaluation requires knowledge of the background incidence rate—or base rate—of myelitis in the general population. Accordingly, in a critical Step 4, we introduced base rate information (1% incidence of myelitis), allowing participants to compare the number of observed cases with what would be expected by chance alone. We hypothesized that this information would reduce concerns about a causal link between the vaccine and myelitis in the Low—but not High—Frequency condition, where the observed number of cases (10 of 1,000 trial participants) matched the population base rate (1%). Evaluating whether observed cases exceed expectations requires recognizing the relevance of base rate information and translating it into a comparable format (e.g., converting a 1% incidence rate into an expectation of 10 cases per 1,000 vaccinated individuals), processes that may be more likely to be performed by individuals higher in numeracy (Kahan et al., 2017; Obrecht & Chesney, 2016). Accordingly, we predicted that numeracy would moderate this effect, such that more numerate individuals would be more likely to incorporate base rate information into their judgments and, in turn, show a greater reduction in perceived risk in the Low Frequency condition relative to less numerate individuals.

Although strong statistical reasoners may recognize that adverse outcomes observed at a rate matching the population base rate do not imply a causal link between the vaccine and the outcome, this inference may not be readily apparent to all. In Step 5, we made the implications of the base rate explicit, stating that, on the basis of the population base rate, approximately 10 out of 1,000 vaccinated individuals would be expected to develop myelitis even in the absence of a causal link. Notably, this step did not introduce new information but instead clarified the

implications of data presented during the previous step (Step 4). As such, participants who recognized these implications should have already incorporated base rate information into their judgments, resulting in minimal change at this step. However, we expected that many participants—particularly those lower in numeracy—would fail to utilize this information in Step 4 and would therefore, in the Low Frequency condition, exhibit a decrease in perceived evidence strength and personal concern in Step 5 when the equivalence between observed and expected cases was made explicit. Thus, we propose that while base rate information can mitigate concerns about adverse outcomes in clinical trials when appropriately interpreted, its implications are often overlooked, particularly among less numerate individuals. Such failures in statistical reasoning may contribute to inflated perceptions of risk following reports of adverse events during clinical trials, even when those events are consistent with chance expectations. Accordingly, the present research tests whether making the implications of base rate information explicit reduces misperceptions, thereby identifying a mechanism underlying vaccine hesitancy and providing a practical strategy for improving public understanding of vaccine safety.

Multiple-Testing and the Interpretation of Adverse Events in Clinical Trials

When evaluating whether an adverse outcome provides evidence for a causal link to vaccination, base rate information is not the only relevant consideration. Clinical trial participants are typically monitored for hundreds or even thousands of potential adverse outcomes, increasing the likelihood that some outcomes will appear more frequently than expected by chance alone. In a final step (Step 6), we introduced this consideration to assess whether drawing attention to the elevated false-positive rate associated with multiple testing would reduce participants' concern, particularly in the High Frequency condition where the observed number of myelitis cases exceeded the population base rate.

Although the consequences of multiple testing are subtle, they can be illustrated with a concrete example. Suppose that, in a sample of 1,000 clinical trial participants, researchers test for the presence of 500 different diseases following vaccination. Assume that each disease occurs with a probability of .01, is independent of the probability of the other diseases, and is unaffected by vaccination. Even though the expected number of cases for each disease is 10 ($1,000 \times .01$), sampling variability ensures that some diseases will occur more frequently and others less frequently. Thus, if we obtain a case count of each disease among the 1,000 trial participants and select the disease, out of 500 diseases, that has the highest count, this value will systematically exceed the base rate expectation. We can demonstrate this by simulating this process, examining what, on average, is the value of the highest case count of the 500 diseases. In doing so, we find that the mean case count for the disease observed most frequently is 20.79 with a standard deviation of 1.52, with additional simulations demonstrating that similar highest case counts are observed when testing for 100 or 1,000 diseases (see Figure 1). In short, when many outcomes are monitored, we should not be surprised to find that one disease has an observed frequency that is twice what is expected based on the population base rate.²

Although these simulations rely on simplifying assumptions, they illustrate a general statistical principle: the more outcomes that are monitored, the greater the likelihood of observing apparently anomalous—and potentially concerning—results. Accordingly, we examined whether making this principle explicit would attenuate participants' concerns about a causal link between the vaccine and myelitis when the observed number of cases exceeded the

² In Experiment 1, the number of observed myelitis cases (25) exceeded the mean highest case count obtained in simulations assuming monitoring of 100, 500, or 1,000 diseases (18.69, 20.79, and 21.62, respectively). As such, even when considering the effects of multiple testing, the observed frequency remained higher than would typically be expected by chance alone, potentially limiting the extent to which this information attenuated concern. In Experiment 2, we addressed this limitation by reducing the number of observed cases (from 25 to 20) to better align with the expected highest case count under conditions involving monitoring of hundreds of adverse outcomes.

population base rate, and whether this effect depended on individuals’ level of numeracy. More broadly, we test whether the statistical realities of multiple testing—and the elevated rates of adverse outcomes it can produce—constitute a barrier to public trust in vaccine safety, and whether making this reality transparent can help reduce unwarranted concern that leads to vaccine hesitancy.

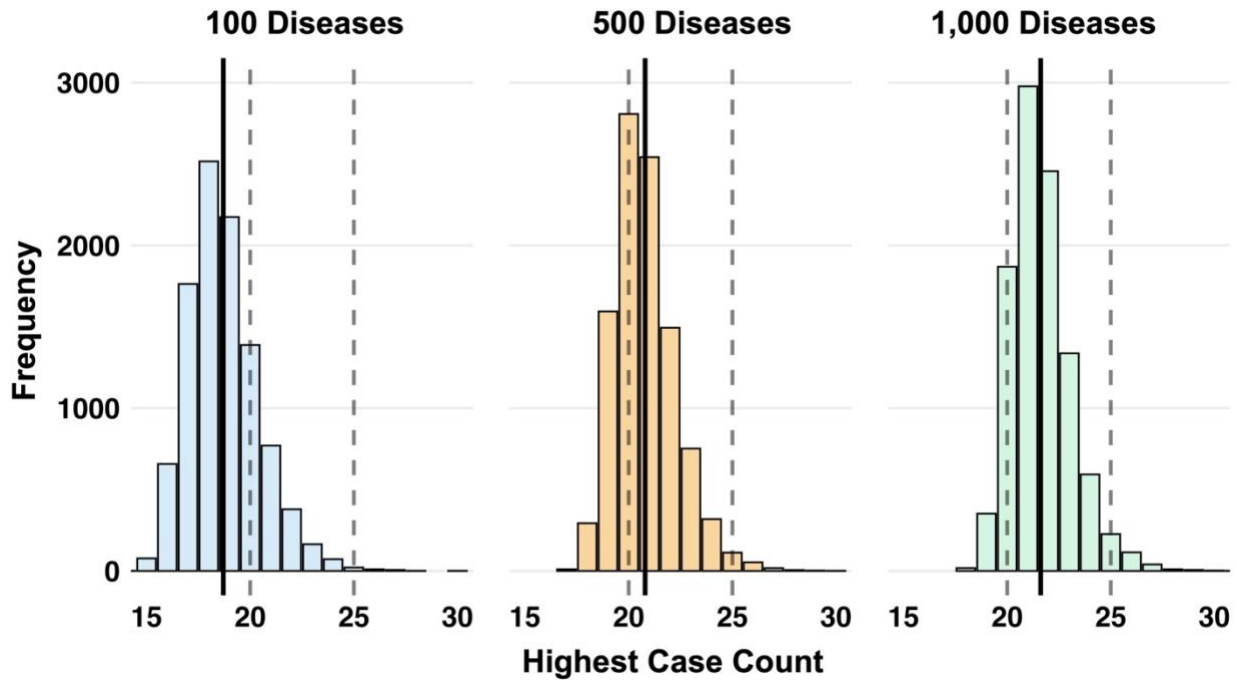


Figure 1 | Clinical Trial Simulations. Distributions of the highest observed case count across diseases in simulated clinical trials. Each panel shows results from 10,000 simulations of a clinical trial with 1,000 participants monitored for 100, 500, or 1,000 diseases (left to right). In all simulations, each disease occurred independently with a probability of .01 and was unaffected by vaccination. Within each simulation, the disease with the highest case count was identified, and the distribution of these highest counts are shown. Solid vertical lines denote the mean highest case count (100 diseases: 18.69; 500 diseases: 20.79; 1,000 diseases: 21.62). Dashed lines mark the number of myelitis cases observed in the experimental vignettes (25 in Experiment 1; 20 in Experiment 2).

Experiment 1

Method

Participants

Three hundred and one participants (43% Female; $M_{\text{age}} = 40.01$, $SD_{\text{age}} = 12.11$; 11% Asian, 4% Black/African American, 5% Hispanic, 79% White/European, 1% Other) were recruited from Amazon Mechanical Turk on March 31st, 2021. We conducted an a priori power analysis using G*Power (Faul et al., 2009) for the hypothesized Illness Frequency by Information Step interaction on critical trials. Assuming a small effect size ($f = 0.10$; $\eta_p^2 \approx .01$) and an alpha level of .05, this analysis indicated that a minimum sample of 266 participants would be required to detect the hypothesized interaction with 90% power. On the basis of this analysis, we pre-registered a sample size of 300 participants. To be eligible to participate in Experiment 1, participants were required to: (1) reside in the United States, (2) have completed 100 or more Human Intelligence Tasks (HITs) on Mechanical Turk, (3) possess a Mechanical Turk submission approval rating greater than or equal to 99%, and (4) correctly respond to two simple questions. Participants received \$1.00 USD upon completion of a 12-minute online questionnaire.

Measures and Materials

Participants in Experiment 1 were asked to consider a scenario in which, during the initial public rollout of a new vaccine, there were confirmed reports of people developing a rare disease (myelitis) following vaccination. Their task was to evaluate the evidence for a causal link between the vaccine and myelitis. Task-relevant information was presented across six sequential steps (see Table 1), with each step adding a new piece of information. During each step, participants responded to multiple items asking them to assess the sufficiency of the information

they received, the strength of the evidence linking the vaccine to myelitis, and their level of concern regarding the vaccine's safety. As such, this design allowed us to examine how participants' evaluations of vaccine safety evolved as new information was accumulated.

Evidence Strength. Participants' perception of the strength of evidence for a causal link between the vaccine and myelitis was assessed using two items. First, participants were asked, "Based on the information that you have, how strong is the evidence that the vaccine causes myelitis?" and responded on a 7-point scale that ranged from 1 (Extremely weak) to 7 (Extremely strong). Next, they were asked "How likely is it that the vaccine causes myelitis?" and indicated their response using a 100-point slider (0-100). Responses to both items were normalized (i.e., converted to a 0-1 score) and averaged to form a composite measure of perceived evidence strength.

Personal Concern. Two questions assessed participants' personal level of concern regarding vaccine safety. That is, participants were asked "How worried would you feel about the safety of the vaccine?" and "To what extent would you feel concerned regarding these reported cases of myelitis?" Responses to both questions were provided on a 5-point scale ranging from 1 (not at all worried/concerned) to 5 (extremely worried/concerned). Responses to both items were normalized and averaged to form a composite measure of personal concern.

Information Sufficiency. At each step, participants were asked: "Do you think you have enough information to judge the strength of the evidence that the vaccine does (or does not) cause myelitis?" Responses to this question were provided on a 5-point scale that ranged from 1 (I do not have any of the information that I need) to 5 (I have all of the information I need).

Numeracy. Following experimental trials, participants completed four Berlin Numeracy Test items (BNT; Cokely et al., 2012) designed to test their level of numeracy. For example, one

item asked “Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an even number (2 or 4)?” All responses were provided within a free-entry text box. We calculated a numeracy score for each participant reflecting the number of correct responses they provided (0-4) to BNT items.

Design and Procedure

On six occasions, participants were presented with additional information pertaining to the rollout of a new vaccine for which adverse events had been observed (see Table 1), and responded to evidence strength, personal concern, and information sufficiency items. We randomly assigned participants to either a Low or High Illness Frequency condition, manipulating whether 10 or 25 people developed myelitis following vaccination in the presented scenario. Thus, Experiment 1 featured a 6 (Information Step: 1, 2, 3, 4, 5, 6; [within-subjects]) by 2 (Illness Frequency: Low, High [between-subjects]) mixed design. Following experimental trials, participants completed four BNT items and responded to three demographic questions. Prior to exiting the study, they were explicitly told that the presented scenario and statistics were entirely fictional.

Statistics and Software

All analyses were conducted in RStudio v2024.12.0+467 (Posit Team, 2024) with R v4.4.2 (R Core Team, 2024). Linear mixed-effects models were fitted to participant data using the lme4 package v1.1.35.5 with degrees of freedom estimated using the Satterthwaite method (Bates et al., 2015; Kuznetsova et al., 2017). Plots were generated using ggplot2 (v3.5.1; Wickham, 2016).

Transparency and Openness

For both experiments, we collected the full sample prior to data analyses and report how we determined our sample size, all data exclusions, all manipulations, and all measures used. All measures and materials presented within Experiments 1 and 2 can be viewed in the supplementary materials (Part A). All data and analysis scripts have been made publicly available (<https://osf.io/4jw3m/>). The hypotheses, methods, and analysis plan for Experiments 1 and 2 were preregistered through Open Science Framework. These pre-registrations can be viewed via the following links (Exp 1: <https://osf.io/7fm49/>; Exp 2: <https://osf.io/tvzmn/>). There were minor deviations from each preregistration (see Supplementary Materials Part B). All experiments were reviewed and received ethics clearance from a University of Waterloo Research Ethics Committee (ORE #42866). All participants provided informed consent prior to their participation.

Results and Discussion

Assessing the likelihood of a causal link between the vaccine and myelitis requires more than knowledge of the number of cases observed in the clinical trial; it also depends on the base rate of myelitis in the general population. At a critical fourth step, participants were provided with base rate information, enabling comparison between the number of cases observed in the clinical trial (10 or 25 out of 1,000, depending on condition) and the number expected by chance alone (10 out of 1,000). We hypothesized that this information would amplify the influence of Illness Frequency on participants' judgments. Consistent with this prediction, we observed a significant Illness Frequency by Information Step (Step 3 vs. Step 4) interaction (see Table 2; Figure 2). The introduction of base rate information reduced perceptions of a causal link between the vaccine and myelitis, with this reduction being larger in the Low Frequency condition, $\Delta M =$

-0.12, 95% CI [-0.15, -0.09], $t(299) = -8.63, p < .001$, than High Frequency condition, $\Delta M = -0.03$, 95% CI [-0.06, -0.00], $t(299) = -2.03, p = .043$. The same pattern emerged for judgments of personal concern (see Table 2). Thus, providing base rate information attenuated perceptions of a causal link between the vaccine and myelitis and reduced concerns about vaccine safety—particularly when the observed number of cases did not exceed the population base rate.

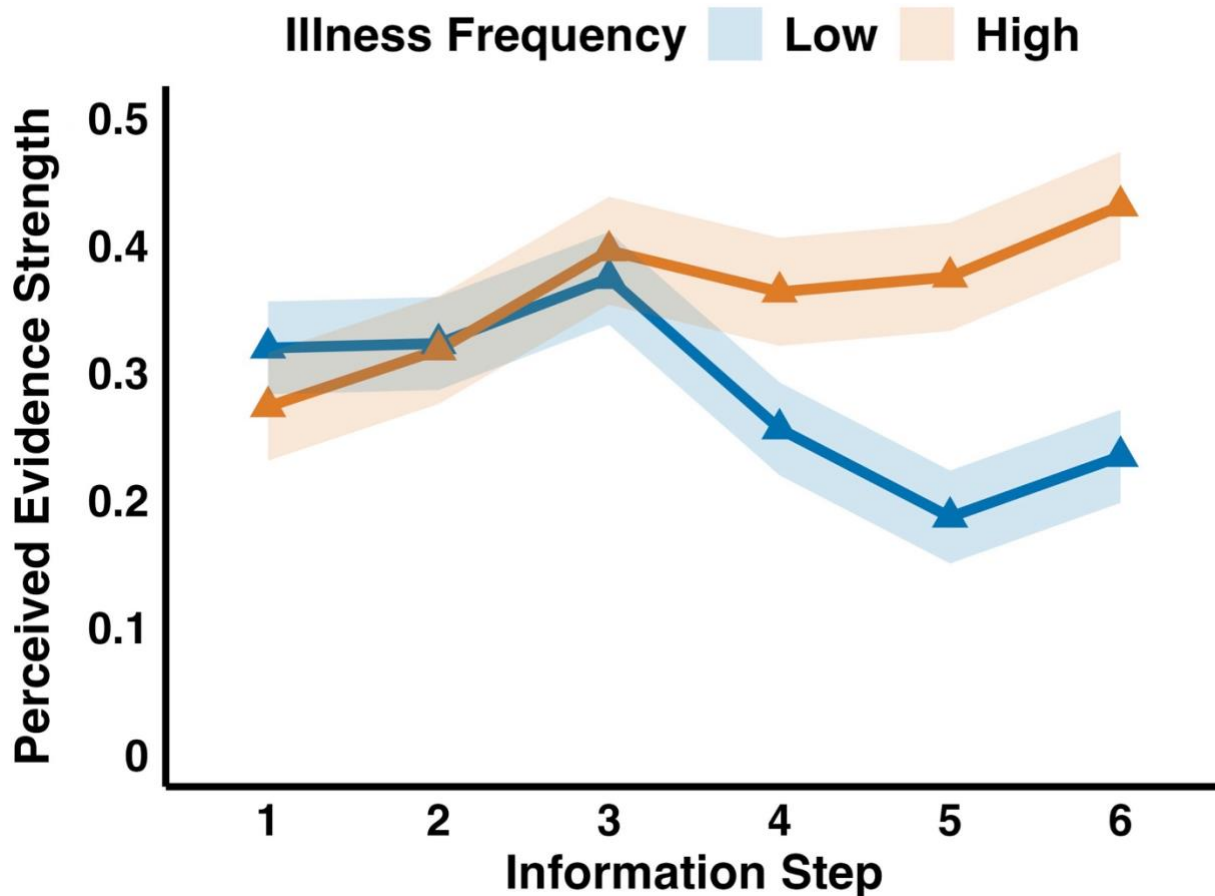


Figure 2 | Experiment 1: Perceived Evidence Strength as a Function of Illness Frequency and Information Step. Figure displays model-estimated marginal means from a linear mixed-effects model predicting perceived evidence strength from Illness Frequency condition (Low vs. High), Information Step, and their interaction. The model included a random intercept for participant to account for repeated observations. Points represent estimated marginal means derived from the fitted model, with error bands reflecting the 95% confidence intervals around these means. Brief description of each Information Step: 1 = Reports of myelitis following vaccination; 2 = Information about the number of cases observed; 3 = Information about the number of people vaccinated; 4 = Presentation of the population base rate; 5 = Calculation of the

number of myelitis cases expected by chance; 6 = Consideration of monitoring for many adverse outcomes.

Table 2 | Illness Frequency Moderates the Effect of Base Rate Information on Vaccine-Related Judgments (Experiment 1)

Variable	Estimate	95% CI	<i>t</i>	df	<i>p</i>
Evidence Strength (DV)					
Intercept	0.37	[0.34, 0.41]	19.71	386	<.001
Illness Frequency	0.02	[-0.04, 0.08]	0.75	386	.457
Information Step	-0.12	[-0.14, -0.09]	-8.63	299	<.001
Frequency x Step	0.09	[0.04, 0.13]	4.09	299	<.001
Personal Concern (DV)					
Intercept	0.50	[0.45, 0.54]	20.95	362	<.001
Illness Frequency	0.02	[-0.05, 0.09]	0.57	362	.570
Information Step	-0.15	[-0.18, -0.12]	-10.42	299	<.001
Frequency x Step	0.09	[0.05, 0.14]	4.08	299	<.001

Note. Results of linear mixed-effects models predicting judgments of evidence strength and personal concern from Illness Frequency condition (0 = Low Frequency, 1 = High Frequency), Information Step (0 = Step 3, 1 = Step 4) and their interaction (Frequency x Step). Both models included a random intercept for participant to account for repeated observations. Evidence Strength Model: $R_m^2 = .05$, $R_c^2 = .75$; Personal Concern Model: $R_m^2 = .05$, $R_c^2 = .82$.

In the subsequent step (Step 5), we made explicit that, given the base rate information provided in Step 4, 10 out of every 1,000 vaccinated individuals would be expected to develop myelitis by chance alone (i.e., in the absence of a causal link between the vaccine and myelitis). We examined whether making this expectation explicit further amplified the impact of the Illness Frequency manipulation. Consistent with this prediction, we observed a significant Illness Frequency by Information Step (Step 4 vs. Step 5) interaction (see Table 3). In the Low Frequency condition—where the observed number of myelitis cases matched the expected base rate—participants perceived weaker evidence for a causal link following this clarification, $\Delta M = -0.07$, 95% CI [-0.09, -0.05], $t(299) = -7.25$, $p < .001$. In contrast, perceptions of evidence

strength did not decrease in the High Frequency condition, where the observed number of cases exceeded that expected by chance, $\Delta M = 0.01$, 95% *CI* [-0.01, 0.03], $t(299) = 1.06$, $p = .290$. The same pattern was once again observed for judgments of personal concern. Therefore, even without introducing new information, making the implications of the base rate explicit reduced perceptions of a causal link and tempered concerns about vaccine safety when observed outcomes matched population expectations, suggesting that participants did not make full use of base rate information in the previous step.

Table 3 | Illness Frequency Moderates the Effect of Base Rate Clarification on Vaccine-Related Judgments (Experiment 1)

Variable	Estimate	95% CI	<i>t</i>	df	<i>p</i>
Evidence Strength (DV)					
Intercept	0.26	[0.22, 0.29]	13.28	338	<.001
Illness Frequency	0.11	[0.05, 0.17]	3.64	338	<.001
Information Step	-0.07	[-0.09, -0.05]	-7.25	299	<.001
Frequency x Step	0.08	[0.05, 0.11]	5.53	299	<.001
Personal Concern (DV)					
Intercept	0.34	[0.30, 0.39]	14.37	336	<.001
Illness Frequency	0.11	[0.04, 0.19]	3.08	336	.002
Information Step	-0.06	[-0.09, -0.04]	-5.45	299	<.001
Frequency x Step	0.05	[0.02, 0.09]	3.00	299	.003

Note. Results of linear mixed-effects models predicting judgments of evidence strength and personal concern from Illness Frequency condition (0 = Low Frequency, 1 = High Frequency), Information Step (0 = Step 4, 1 = Step 5) and their interaction (Frequency x Step). Both models included a random intercept for participant to account for repeated observations. Evidence Strength Model: $R_m^2 = .09$, $R_c^2 = .89$; Personal Concern Model: $R_m^2 = .05$, $R_c^2 = .89$.

In a final step, we introduced an additional consideration that is often overlooked when evaluating evidence for a causal link between a vaccine and an adverse outcome: clinical trial participants are monitored for hundreds of adverse outcomes, increasing the likelihood that some

will appear more frequently in the vaccinated group purely by chance. We hypothesized that making this consideration explicit would attenuate the effect of Illness Frequency by reducing perceptions of evidence strength, particularly in the High Frequency condition. This prediction was not supported. Instead, evidence strength judgments increased following this information (Low Frequency condition: $\Delta M = 0.05$, 95% *CI* [0.03, 0.07], $t(299) = 5.31$, $p < .001$; High Frequency condition: $\Delta M = 0.06$, 95% *CI* [0.04, 0.08], $t(299) = 5.34$, $p < .001$), and the predicted Illness Frequency by Information Step (Step 5 vs. Step 6) interaction was not observed, $b = 0.01$, 95% *CI* [-0.02, 0.04], $t(299) = 0.59$, $p = .558$. A similar pattern was observed for judgments of personal concern. Thus, highlighting the role of multiple testing did not mitigate perceptions of a causal link between the vaccine and myelitis or concerns about vaccine safety, regardless of condition.

Numeracy Moderates Sensitivity to Base Rate Information

Given that the task relied heavily on statistical reasoning, we examined whether individual differences in numeracy moderated participants' responses to key pieces of information. First, we tested whether numeracy shaped responses following the introduction of base rate information in Step 4, reasoning that more numerate individuals would be better able to recognize the relevance of base rate information and properly use it to inform their judgments. Consistent with this account, we observed a significant Illness Frequency by Information Step (Step 3 vs. Step 4) by Numeracy interaction for evidence strength judgments, $b = 0.05$, 95% *CI* [0.02, 0.08], $t(297) = 3.28$, $p = .001$. We split numeracy at the mean ($M = 1.61$) to probe this 3-way interaction. Follow-up analyses revealed that the effect of Illness Frequency was amplified among more numerate participants following the presentation of base rate information, $\Delta M = 0.14$, 95% *CI* [0.08, 0.20], $t(297) = 4.54$, $p < .001$, but not among those lower in numeracy, $\Delta M =$

0.04, 95% *CI* [-0.02, 0.09], $t(297) = 1.40$, $p = .162$ (Figure 3). A similar pattern was observed for judgments of personal concern. Thus, base rate information reduced perceptions of a causal link between the vaccine and myelitis when (1) the number of observed cases did not exceed population expectations and (2) individuals possessed sufficient competency in numerical reasoning. The moderating effect of numeracy was due, at least in part, to more numerate individuals being better able to recognize the relevance of base rate information, as information sufficiency judgments increased more sharply following the introduction of base rate information for more numerate participants, $b = 0.14$, 95% *CI* [0.06, 0.22], $t(299) = 3.30$, $p = .001$.

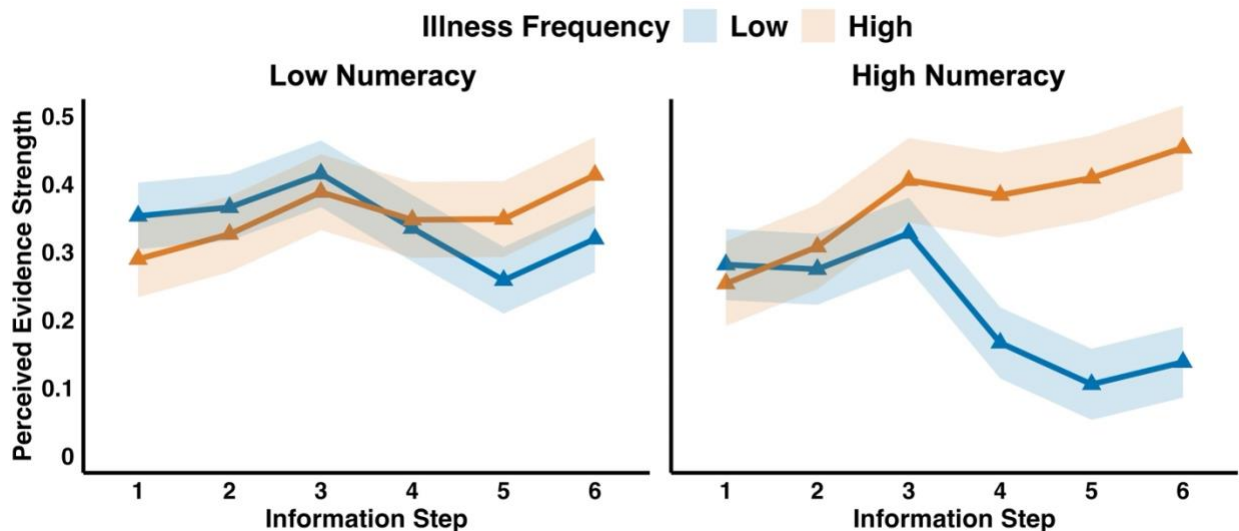


Figure 3 | Experiment 1: Perceived Evidence Strength as a Function of Illness Frequency, Information Step, and Numeracy. Figure displays model-estimated marginal means from a linear mixed-effects model including Illness Frequency condition (Low vs. High), Information Step, and Numeracy (Low vs. High), along with their interaction, as predictors. The model included a random intercept for participant to account for repeated observations. Panels depict estimated marginal means separately for low- and high-numeracy participants. Points represent estimated marginal means and error bands reflect 95% confidence intervals. Brief description of each Information Step: 1 = Reports of myelitis following vaccination; 2 = Information about the number of cases observed; 3 = Information about the number of people vaccinated; 4 = Presentation of the population base rate; 5 = Calculation of the number of myelitis cases expected by chance; 6 = Consideration of monitoring for many adverse outcomes.

We did not observe comparable three-way interactions for later steps (Step 4 vs. Step 5 or Step 5 vs. Step 6; $ps > .203$). Thus, although more numerate individuals reduced their concern when adverse outcomes aligned with base rate expectations, explicitly highlighting the implications of base rate information (Step 5) did not disproportionately influence the judgments of less numerate participants, who were less responsive to base rate information. Moreover, participants did not mitigate their concern in response to information highlighting the role of multiple testing, regardless of numeracy level or Illness Frequency condition. Thus, even individuals high in numeracy may fail to incorporate multiple testing considerations into their judgments, highlighting a potential barrier to public trust in vaccine safety.

Experiment 2

Highlighting the role of multiple testing in Experiment 1 did not reduce perceptions of a causal link between the vaccine and myelitis or concerns about vaccine safety, regardless of participants' level of numeracy. One explanation is that, in the High Frequency condition—where such information should be most impactful—the number of observed cases remained higher than what might reasonably be expected by chance even when accounting for multiple testing, thereby limiting the effectiveness of this consideration. To address this limitation in Experiment 2, we reduced the number of observed myelitis cases in the High Frequency condition to better align with simulation-based estimates of the highest case count expected when monitoring for hundreds of potential adverse outcomes.

Additionally, although participants in Experiment 1 were informed that myelitis was one of many outcomes monitored, the link between multiple testing and the increased likelihood of false positives was not made explicit. Accordingly, we revised Step 6 information in Experiment 2 to make this connection explicit. Beyond these modifications, Experiment 2 closely mirrored

the design of Experiment 1, allowing us to replicate key findings while testing whether clearer communication about multiple testing and false positives could reduce concerns about vaccine safety, particularly when observed cases of myelitis exceed the population base rate.

Method

Participants

Three hundred and one U.S. residents (47% Female; $M_{\text{age}} = 43.53$, $SD_{\text{age}} = 12.45$; 4% Asian, 7% Black/African American, 3% Hispanic, 86% White/European, 1% Other; 78% College Educated) were recruited from Amazon Mechanical Turk on May 4th, 2021. Experiment 2 used the same recruitment criteria as Experiment 1. Participants received \$1.80 USD upon completion of a 15-minute online questionnaire.

Measures and Materials

Experiment 2 used the same study materials as Experiment 1, with two exceptions. First, we modified the information presented on the final experimental trial (Step 6) to more clearly convey the relationship between the number of adverse outcomes monitored and the likelihood of false positives (see Table 4). Second, we reduced the number of myelitis cases observed among clinical trial participants in the High Frequency condition from 25 to 20. Aside from these changes, participants were presented with the same vaccine rollout scenario and sequential information as in Experiment 1 and responded to the same measures of information sufficiency, evidence strength, and personal concern at each step.

Table 4

Experiment 2: Revised Information Step

Step	(Additional) Information
6	<p>Suppose that Myelitis is just one of hundreds of potential rare adverse effects or illnesses (e.g., rashes, arthritis, stroke, encephalitis, asthma, tinnitus) that public health authorities monitored for among the 1,000 vaccinated individuals. For each adverse outcome, as with myelitis, we can compare how often it occurs following vaccination against a "background incidence" of how often it would occur even in the absence of vaccination. If the adverse outcome occurs more often following vaccination than expected by its background incidence, that might suggest the outcome is caused by the vaccine. But chance also plays a role. We would expect a fair coin toss to produce heads 50% of the time, but in any series of coin tosses the actual proportion of heads might be somewhat higher or lower than 50%.</p> <p>The same is true for adverse outcomes: Just by chance, even if they were not caused by the vaccine, some adverse outcomes will occur more frequently, and others less frequently, than would be expected based on their background incidence. If we monitor for a large number of adverse outcomes, then, there are bound to be a few "false positives," meaning a few outcomes that occur more frequently than expected (like getting more than 50% heads in a series of coin tosses), just due to chance rather than to any consequence of vaccination.</p>

Note. The final information step (Step 6) was revised to provide a more detailed description of the consequences of multiple testing for false positives relative to Experiment 1. The preceding five steps were identical to those used in Experiment 1 and are reported in Table 1 and the supplementary materials (Part A).

Design and Procedure

The design and procedure of Experiment 2 closely mirrored that of Experiment 1.

However, Experiment 2 included additional measures administered after experimental trials.

Specifically, participants completed an additional scale designed to test their level of numeracy

(8-item Rasch-based Numeracy Scale; Weller et al., 2013)³ and responded to additional

³ Analyses using this alternative measure are reported exclusively in the supplemental materials and replicate all results reported in the main text (which used the Berlin Numeracy Test to measure participant numeracy), with one exception: the Illness Frequency by Information Step (Step 3 vs. Step 4) by Numeracy interaction, while once again observed for judgments of perceived evidence strength, $b = 0.04$, 95% *CI* [0.02, 0.07], $t(297) = 3.61$, $p < .001$, was not observed for judgments of personal concern, $b = 0.02$, 95% *CI* [-0.01, 0.05], $t(297) = 1.61$, $p = .108$.

demographic questions assessing their level of education, political ideology, and beliefs about the general safety of vaccines.

Results and Discussion

Before testing whether reducing the number of myelitis cases in the High Frequency condition and making the link between multiple testing and false positives more explicit influenced participants' judgments, we first sought to replicate key findings from Experiment 1. To this end, we again examined whether introducing base rate information (Step 4)—which enabled comparison between the observed number of myelitis cases (10 or 20 out of 1,000, depending on condition) and the number expected by chance alone (10 out of 1,000)—amplified the influence of Illness Frequency on participants' judgments. As in Experiment 1, we observed an Illness Frequency by Information Step (Step 3 vs. Step 4) interaction (see Table 5; Figure 4). Base rate information reduced perceptions of a causal link between the vaccine and myelitis, with this reduction being more pronounced in the Low Frequency, $\Delta M = -0.17$, 95% *CI* [-0.20, -0.13], $t(299) = -10.17$, $p < .001$, compared to High Frequency condition, $\Delta M = -0.05$, 95% *CI* [-0.08, -0.02], $t(299) = -3.13$, $p = .002$. The same pattern was observed for judgments of personal concern (see Table 5). Therefore, consistent with Experiment 1, base rate information decreased perceptions of a causal link between the vaccine and myelitis and reduced concerns about vaccine safety—particularly when the number of myelitis cases aligned with chance expectations.

Table 5 | Illness Frequency Moderates the Effect of Base Rate Information on Vaccine-Related Judgments (Experiment 2)

Variable	Estimate	95% CI	<i>t</i>	df	<i>p</i>
Evidence Strength (DV)					
Intercept	0.36	[0.32, 0.39]	18.03	417	<.001
Illness Frequency	0.05	[-0.00, 0.10]	1.83	417	.068
Information Step	-0.17	[-0.20, -0.13]	-10.17	299	<.001
Frequency x Step	0.12	[0.07, 0.16]	5.03	299	<.001
Personal Concern (DV)					
Intercept	0.43	[0.39, 0.48]	18.60	391	<.001
Illness Frequency	0.07	[0.01, 0.14]	2.22	391	.027
Information Step	-0.18	[-0.21, -0.15]	-10.39	299	<.001
Frequency x Step	0.11	[0.07, 0.16]	4.71	299	<.001

Note. Results of linear mixed-effects models predicting judgments of evidence strength and personal concern from Illness Frequency condition (0 = Low Frequency, 1 = High Frequency), Information Step (0 = Step 3, 1 = Step 4) and their interaction (Frequency x Step). Both models included a random intercept for participant to account for repeated observations. Evidence Strength Model: $R_m^2 = .10$, $R_c^2 = .69$; Personal Concern Model: $R_m^2 = .10$, $R_c^2 = .75$.

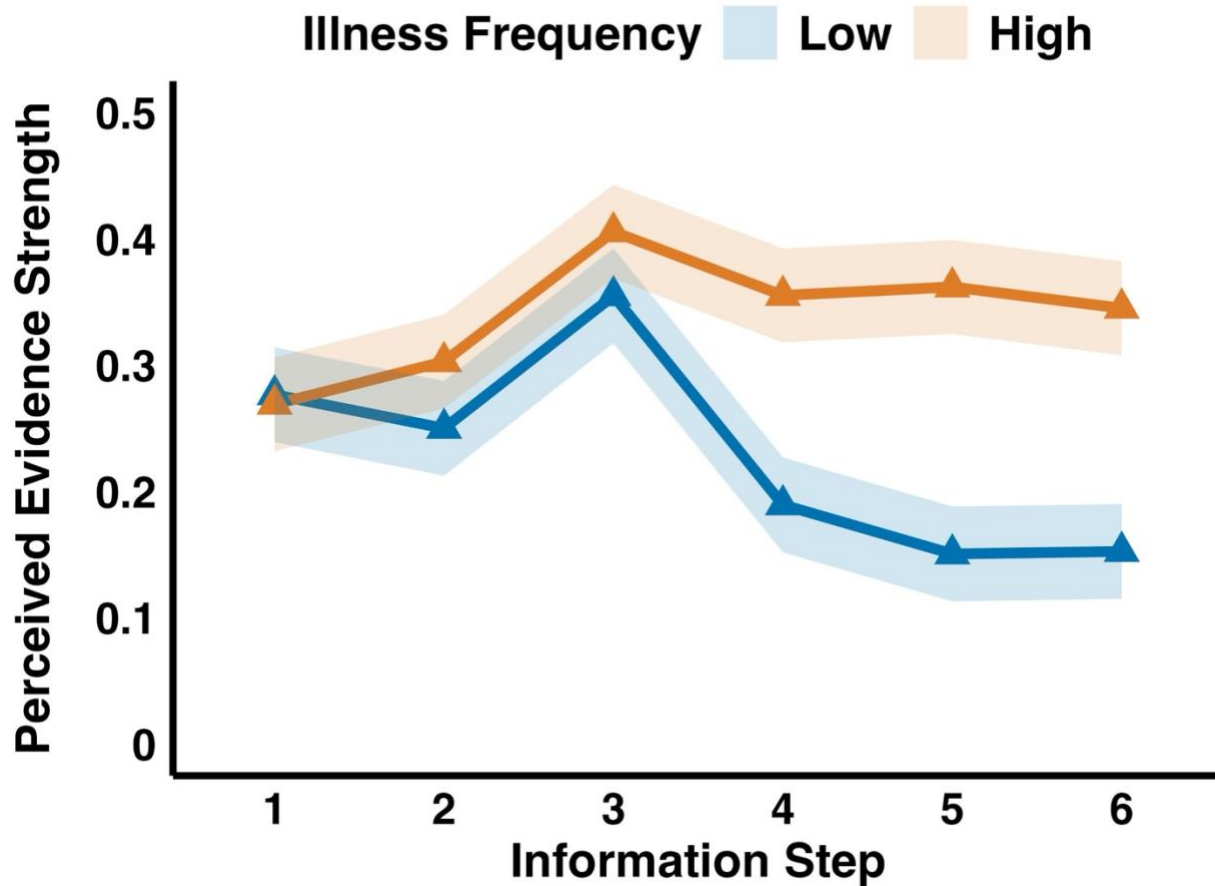


Figure 4 | Experiment 2: Perceived Evidence Strength as a Function of Illness Frequency and Information Step. Figure displays model-estimated marginal means from a linear mixed-effects model predicting perceived evidence strength from Illness Frequency condition (Low vs. High), Information Step, and their interaction. The model included a random intercept for participant to account for repeated observations. Points represent estimated marginal means derived from the fitted model, with error bands reflecting the 95% confidence intervals around these means. Brief description of each Information Step: 1 = Reports of myelitis following vaccination; 2 = Information about the number of cases observed; 3 = Information about the number of people vaccinated; 4 = Presentation of the population base rate; 5 = Calculation of the number of myelitis cases expected by chance; 6 = Consideration of monitoring for many adverse outcomes.

The following step (Step 5) made the implications of the provided base rate information explicit, stating that, given the population base rate, 10 out of every 1,000 vaccinated individuals would be expected to develop myelitis during clinical trials, even if the vaccine does not cause

myelitis. As in Experiment 1, this clarification increased the impact of Illness Frequency, resulting in the hypothesized Illness Frequency by Information Step (Step 4 vs. Step 5) interaction (see Table 6). In the Low Frequency condition—where observed cases were consistent with base rate expectations—participants judged the evidence for a causal link as weaker following this clarification, $\Delta M = -0.04$, 95% *CI* [-0.06, -0.02], $t(299) = -4.66$, $p < .001$. Conversely, in the High Frequency condition, where the number of myelitis cases exceeded chance expectations, evidence strength judgments were statistically unchanged, $\Delta M = 0.01$, 95% *CI* [-0.01, 0.02], $t(299) = 0.81$, $p = .419$. A similar pattern emerged for judgments of personal concern, although the hypothesized Illness Frequency by Information Step interaction was not statistically significant ($p = .087$; see Table 6). Nonetheless, judgments of evidence strength again indicate that individuals may often fail to recognize the implications of base rate information, as simply making these implications explicit reduced evidence strength judgments when the observed number of cases aligned with the population base rate.

Table 6 | Illness Frequency Moderates the Effect of Base Rate Clarification on Vaccine-Related Judgments (Experiment 2)

Variable	Estimate	95% CI	<i>t</i>	df	<i>p</i>
Evidence Strength (DV)					
Intercept	0.19	[0.15, 0.23]	9.67	327	<.001
Illness Frequency	0.17	[0.11, 0.22]	6.02	327	<.001
Information Step	-0.04	[-0.06, -0.02]	-4.66	299	<.001
Frequency x Step	0.05	[0.02, 0.07]	3.88	299	<.001
Personal Concern (DV)					
Intercept	0.26	[0.21, 0.30]	11.18	331	<.001
Illness Frequency	0.19	[0.12, 0.25]	5.84	331	<.001
Information Step	-0.05	[-0.07, -0.03]	-4.98	299	<.001
Frequency x Step	0.02	[-0.00, 0.05]	1.72	299	.087

Note. Results of linear mixed-effects models predicting judgments of evidence strength and personal concern from Illness Frequency condition (0 = Low Frequency, 1 = High Frequency), Information Step (0 = Step 4, 1 = Step 5) and their interaction (Frequency x Step). Both models included a random intercept for participant to account for repeated observations. Evidence Strength Model: $R_m^2 = .14$, $R_c^2 = .92$; Personal Concern Model: $R_m^2 = .12$, $R_c^2 = .91$.

While participants in the Low Frequency condition reduced their perceptions of a causal link between the vaccine and myelitis in response to base rate information, this was not the case in the High Frequency condition, where observed cases exceeded population expectations. Yet, as we have discussed, because clinical trial participants are monitored for hundreds of adverse outcomes, some outcomes will inevitably appear more frequently than expected on the basis of population base rates, even in the absence of any causal effect of the vaccine. We again assessed whether making this consideration explicit would reduce judgments of evidence strength and personal concern, particularly among participants in the High Frequency condition.

The number of myelitis cases in the High Frequency condition was adjusted downward in Experiment 2 (i.e., from 25 to 20) to better correspond to the most frequently observed outcome expected by chance alone in a clinical trial in which participants are monitored for 500 adverse outcomes (10,000 simulations; Simulated Highest Case Count: $M = 20.79$, $SD = 1.52$, see Figure 1). Unlike in Experiment 1, participants in the High Frequency condition did perceive weaker evidence for a causal link between the vaccine and myelitis following this information, $\Delta M = -0.02$, 95% $CI [-0.03, -0.00]$, $t(299) = -2.24$, $p = .026$. However, the hypothesized Illness Frequency by Information Step (Step 5 vs. Step 6) interaction was not statistically significant, $b = -0.02$, 95% $CI [-0.04, 0.00]$, $t(299) = -1.76$, $p = .079$. The same pattern was observed for judgments of personal concern. Thus, making explicit the role of multiple testing and its contribution to false positives produced only a small reduction in perceived evidence strength and concerns about vaccine safety, and was not reliably moderated by Illness Frequency.

Numeracy Moderates Sensitivity to Base Rate Information

We again assessed whether participants' numerical abilities shaped how they responded to key pieces of information. Replicating Experiment 1, more numerate individuals were better able to recognize the relevance of base rate information and incorporate it into their judgments, as indicated by a significant Illness Frequency by Information Step (Step 3 vs. Step 4) by Numeracy interaction for evidence strength judgments, $b = 0.07$, 95% $CI [0.03, 0.10]$, $t(297) = 3.97$, $p < .001$. To probe this interaction, we again split numeracy at the mean ($M = 1.55$). Follow-up analyses showed that the effect of Illness Frequency became more pronounced after the introduction of base rate information among participants higher in numeracy, $\Delta M = 0.19$, 95% $CI [0.13, 0.26]$, $t(297) = 5.68$, $p < .001$, but not among less numerate participants, $\Delta M = 0.05$, 95% $CI [-0.01, 0.11]$, $t(297) = 1.64$, $p = .102$ (Figure 5). Within the Low Frequency condition, more numerate participants exhibited a larger decrease in perceived evidence strength upon learning that the number of observed cases matched the population base rate, $b = -0.07$, 95% $CI [-0.10, -0.05]$, $t(147) = -5.48$, $p < .001$, indicating greater sensitivity to base rate information. The same pattern was observed for judgments of personal concern. Additionally, information sufficiency judgments increased more sharply following the presentation of base rate information among more numerate participants, $b = 0.09$, 95% $CI [0.00, 0.18]$, $t(299) = 2.03$, $p = .043$, consistent with them being more likely to recognize the relevance of this information.

We did not observe comparable three-way interactions at later steps (Step 4 vs. Step 5 or Step 5 vs. Step 6; $ps > .172$). Thus, making the implications of base rate information explicit (Step 5) reduced concerns about a causal link between the vaccine and myelitis in the Low Frequency condition, regardless of participants' level of numeracy. Furthermore, although the revised multiple testing explanation and reduced number of cases in the High Frequency

condition produced a modest reduction in judgments of evidence strength, this effect was not moderated by numeracy. As such, the statistical realities of multiple testing represent an important barrier to perceptions of vaccine safety—one that is not easily overcome through explanation and is not limited to individuals with lower levels of statistical reasoning.

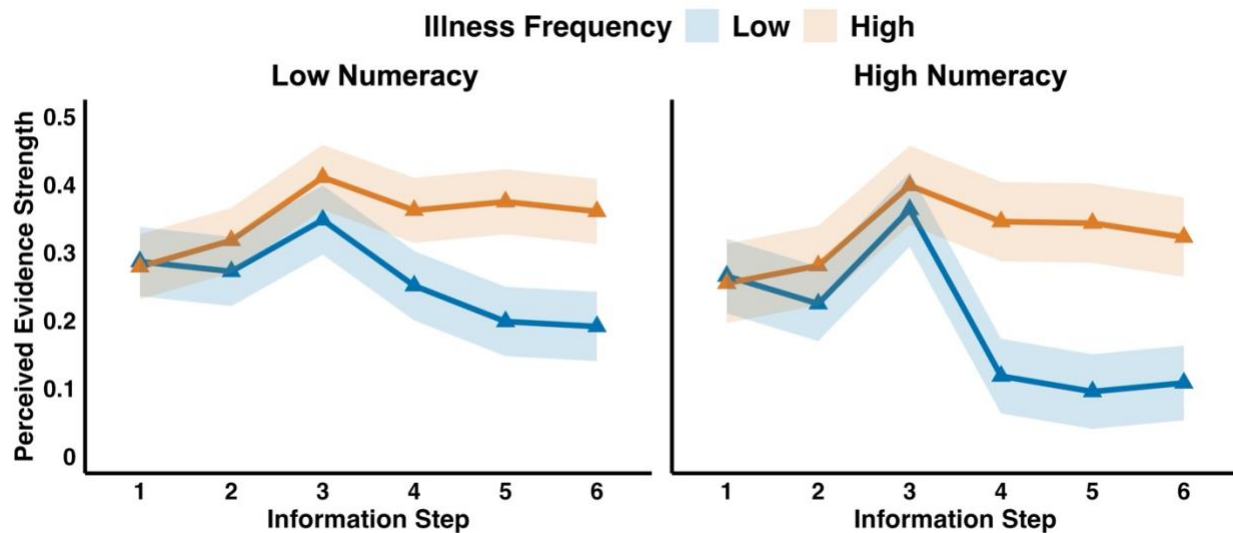


Figure 5 | Experiment 2: Perceived Evidence Strength as a Function of Illness Frequency, Information Step, and Numeracy. Figure displays model-estimated marginal means from a linear mixed-effects model including Illness Frequency condition (Low vs. High), Information Step, and Numeracy (Low vs. High), along with their interaction, as predictors. The model included a random intercept for participant to account for repeated observations. Panels depict estimated marginal means separately for low- and high-numeracy participants. Points represent estimated marginal means and error bands reflect 95% confidence intervals. Brief description of each Information Step: 1 = Reports of myelitis following vaccination; 2 = Information about the number of cases observed; 3 = Information about the number of people vaccinated; 4 = Presentation of the population base rate; 5 = Calculation of the number of myelitis cases expected by chance; 6 = Consideration of monitoring for many adverse outcomes.

General Discussion

The statistical realities of large-scale vaccine trials ensure that adverse events will occur, even when not caused by vaccination. As such, it is critical to understand how members of the public interpret evidence from clinical trials when adverse outcomes are observed, as reports of adverse events may spark concerns about vaccine safety and contribute to vaccine hesitancy,

with significant public health costs (Jia et al., 2023; Lo & Hotez, 2017; Salmon et al., 2015). In the present research, participants expressed concern about a causal link between a vaccine and myelitis upon learning that cases of myelitis had been observed following vaccination (Step 1), despite lacking the information needed to interpret these events. Thus, even in the absence of contextual information, individuals may treat the mere occurrence of adverse events as cause for alarm, underscoring the importance of public health communications that facilitate appropriate contextualization of such events to mitigate unwarranted concern.

Across two experiments, we presented participants with sequential information about a clinical trial in which cases of myelitis were observed following vaccination. In both experiments, providing base rate information reduced perceptions of a causal link between the vaccine and myelitis and attenuated personal concern when the observed number of cases matched population expectations. Critically, these effects depended on participants' ability to recognize the implications of base rate information: more numerate individuals were more responsive to this information, exhibiting larger decreases in perceived risk and greater increases in perceived information sufficiency.

At the same time, results suggest that many individuals failed to incorporate base rate information into their judgments. In the Low Frequency condition of both experiments, making the implications of this information explicit further reduced judgments of evidence strength and personal concern, despite no new information being introduced. Notably, unlike the initial presentation of base rate information, this clarification reduced concern regardless of participants' level of numeracy. Together, these findings highlight a key barrier to effective risk communication: even when relevant statistical information is available, its impact depends on individuals' ability to extract and apply its implications. They also point to a practical solution.

Translating base rate information into concrete expectations (i.e., expected case counts) helps individuals—including those lower in numeracy—appropriately contextualize adverse events, thereby reducing concern when observed frequencies align with chance expectations.

In contrast, we find limited evidence that explaining the link between multiple testing and false positives—where monitoring many outcomes increases the likelihood that at least one will appear elevated by chance—reduces concerns about vaccine safety. Clinical trials monitor participants for many potential adverse outcomes, making it likely that some will appear more frequently in the trial sample than in the general population even when not caused by vaccination. However, highlighting this consideration did not reliably attenuate judgments of evidence strength or personal concern. In Experiment 1, introducing this consideration *increased* perceived evidence strength and concern. In Experiment 2, despite providing a more detailed explanation and ensuring closer alignment between observed frequencies and simulation-based expectations, we observed only small and inconsistent decreases in these judgments. Moreover, responses to this information did not vary as a function of numeracy. Thus, even individuals relatively adept at statistical reasoning seemingly failed to incorporate the implications of multiple testing into their judgments. The idea that monitoring hundreds of adverse outcomes increases the likelihood that at least one will meaningfully exceed its expected frequency (i.e., based on the population base rate) may be difficult to internalize, conflicting with intuitive interpretations of observed frequencies. As such, the statistical consequences of multiple testing represent a persistent obstacle to accurate inference and public perceptions of vaccine safety.

Implications for Public Health Communication

The present research provides practical insights for improving public health communication, revealing that reports of adverse events following vaccination should not only

include base rate information but also explicitly translate this information into expected case counts to improve risk assessments. This relatively simple intervention reduced concerns about vaccine safety when the frequency of myelitis cases aligned with population expectations, even among less numerate individuals. Thus, while prior work has documented individuals' neglect and misuse of base rate information (Bar-Hillel, 1980; Kahneman & Tversky, 1973), the present research demonstrates the consequences of these errors in the public health domain while identifying an effective strategy for communicating base rates in a way that reduces unwarranted concerns about vaccine safety.

Nevertheless, our findings indicate that not all statistical principles are equally tractable. Whereas the implications of base rate information can be made accessible through explicit clarification, other considerations—such as the influence of multiple testing on adverse outcomes—appear less responsive to explanation. Situations in which individuals must account for multiple testing to appropriately contextualize adverse outcomes may therefore benefit from alternative communication strategies. Prior work demonstrates that statistical inference can be improved by presenting complex information in intuitive graphical formats (Galesic et al., 2009; Garcia-Retamero & Cokely, 2013, 2017; Walker et al., 2019, 2023; Zikmund-Fisher et al., 2008). Thus, a promising direction for future research is to examine whether graphical displays can facilitate this inference, for example by depicting the distribution of adverse outcome counts when (a) many outcomes are simultaneously monitored and (b) vaccination has no effect. More broadly, identifying communication strategies that help individuals appropriately contextualize adverse events that appear elevated relative to baseline expectations remains an important challenge for improving public health communication and perceptions of vaccine safety.

Conclusion

Concerns about vaccine safety are not an unavoidable consequence of adverse events following vaccination; rather, they are shaped by how such events are explained and interpreted. The present research shows how individuals' numerical abilities interact with the accumulation and clarification of statistical information to shape vaccine-related judgments. Understanding how people with varying levels of statistical reasoning interpret public health information is essential for improving communication and supporting informed decision-making in public health contexts.

Constraints on Generality

The present findings were observed in samples of American participants recruited via the online crowdsourcing platform Amazon Mechanical Turk during the COVID-19 pandemic (March-May 2021), raising potential questions about generalizability across populations, sampling methods, and contexts. With respect to participants, Americans may differ from those in other countries in their attitudes toward vaccines (Larson et al., 2016), trust in public health institutions (Moucheraud et al., 2021), and exposure to health information (Link et al., 2022; Mede et al., 2026). However, the core processes examined here—namely, the interpretation of statistical evidence, sensitivity to base rate information, and difficulty incorporating multiple testing considerations—reflect general features of human judgment that have been documented across diverse populations. As such, we expect these effects to generalize beyond the U.S. context, although their generalizability, particularly to non-Western or non-WEIRD (Western, Educated, Industrialized, Rich, and Democratic) populations, remains an open question. The exclusive use of online samples also warrants consideration. Although participants recruited from platforms such as Mechanical Turk may differ from nationally representative samples, they

are generally more representative than traditional undergraduate samples and produce data of comparable quality (Buhrmester et al., 2016; Hauser & Schwarz, 2016; Paolacci et al., 2010). Thus, we do not expect the effects observed here to depend on the mode of data collection. Finally, data were collected during a global pandemic, when vaccine-related risks were particularly salient. Although this context may have heightened engagement or concern, the statistical challenges investigated here are not unique to COVID-19 and apply broadly to the interpretation of adverse events in public health contexts. Consistent with this, we ensured that the vaccine scenario was described in general terms to avoid linking it to COVID-19 vaccinations. We therefore expect these effects to extend across contexts, while acknowledging that their magnitude may vary with the salience of vaccine-related risks in public discourse and the extent to which those risks are perceived as personally relevant.

References

- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica, 44*(3), 211-233.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.
- Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299*(18), 999-1001.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*(1), 25-47.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160.
- Fisher, K. A., Bloomstone, S. J., Walder, J., Crawford, S., Fouayzi, H., & Mazor, K. M. (2020). Attitudes toward a potential SARS-CoV-2 vaccine: A survey of US adults. *Annals of Internal Medicine, 173*(12), 964-973.
- Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology, 28*(2), 210-216.
- Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science, 22*(5), 392-399.

- Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics. *Human Factors, 59*(4), 582-627.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*(2), 53-96.
- Guichon, J., & Franco, E. (2021, January 10). Opinion: We must be on guard for vaccine horror stories. *The Globe and Mail*. <https://www.theglobeandmail.com/opinion/article-we-must-be-on-guard-for-vaccine-horror-stories/>
- Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*(1), 400-407.
- Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine, 73*(5), 538-540.
- Jia, K. M., Hanage, W. P., Lipsitch, M., Johnson, A. G., Amin, A. B., Ali, A. R., ... & Swerdlow, D. L. (2023). Estimated preventable COVID-19-associated deaths due to non-vaccination in the United States. *European Journal of Epidemiology, 38*(11), 1125-1128.
- Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy, 1*(1), 54-86.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237-251.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26.

- Larson, H. J., De Figueiredo, A., Xiaohong, Z., Schulz, W. S., Verger, P., Johnston, I. G., ... & Jones, N. S. (2016). The state of vaccine confidence 2016: Global insights through a 67-country survey. *eBioMedicine*, *12*, 295-301.
- Link, E., Baumann, E., Kreps, G. L., Czerwinski, F., Rosset, M., & Suhr, R. (2022). Expanding the health information national trends survey research program internationally to examine global health communication trends: Comparing health information seeking behaviors in the US and Germany. *Journal of Health Communication*, *27*(8), 545-554.
- Lo, N. C., & Hotez, P. J. (2017). Public health and economic consequences of vaccine hesitancy for measles in the United States. *JAMA Pediatrics*, *171*(9), 887-892.
- Miller, S. G., & Edwards, E. (2020, September 8). Oxford vaccine trial on hold because of potential safety issue. *NBC News*. <https://www.nbcnews.com/health/health-news/oxford-vaccine-trial-hold-due-safety-issue-n1239581>
- Mede, N. G., Cologna, V., Berger, S., Besley, J. C., Brick, C., Joubert, M., ... & Schäfer, M. S. (2026). Public communication about science across 68 countries: Global evidence on how people get information and communicate about science-related matters. Retrieved from: https://osf.io/preprints/osf/xb3ha_v1
- Moucheraud, C., Guo, H., & Macinko, J. (2021). Trust in governments and health workers low globally, influencing attitudes toward health information, vaccines. *Health Affairs*, *40*(8), 1215-1224.
- Neumann-Böhme, S., Varghese, N. E., Sabat, I., Barros, P. P., Brouwer, W., van Exel, J., ... & Stargardt, T. (2020). Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. *The European Journal of Health Economics*, *21*(7), 977-982.

Obrecht, N. A., & Chesney, D. L. (2016). Prompting deliberation increases base-rate use.

Judgment and Decision Making, 11(1), 1-6.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon

Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419.

Posit Team. (2024). *RStudio: Integrated Development Environment for R*. In Posit Software.

<http://www.posit.co/>

R Core Team. (2024). *R: A language and environment for statistical computing*. In R Foundation

for Statistical Computing. <http://www.R-project.org>

Salmon, D. A., Dudley, M. Z., Glanz, J. M., & Omer, S. B. (2015). Vaccine hesitancy: Causes,

consequences, and a call to action. *Vaccine*, 33, D66-D71.

Smith, L. E., Amlôt, R., Weinman, J., Yiend, J., & Rubin, G. J. (2017). A systematic review of

factors affecting vaccine uptake in young children. *Vaccine*, 35(45), 6059-6069.

Solis Arce, J. S., Warren, S. S., Meriggi, N. F., Scacco, A., McMurry, N., Voors, M., ... & Omer,

S. B. (2021). COVID-19 vaccine acceptance and hesitancy in low-and middle-income countries. *Nature Medicine*, 27(8), 1385-1394.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological*

Bulletin, 76(2), 105-110.

Walker, A. C., Stange, M., Dixon, M. J., Koehler, D. J., & Fugelsang, J. A. (2019). Graphical

depiction of statistical information improves gambling-related judgments. *Journal of Gambling Studies*, 35(3), 945-968.

Walker, A. C., Stange, M., Dixon, M. J., Fugelsang, J. A., & Koehler, D. J. (2023). Using icon

arrays to communicate gambling information reduces the appeal of scratch card games.

Journal of Gambling Studies, 39(1), 363-382.

- Walsh, F. (2020, September 9). Coronavirus: Oxford University vaccine trial paused after participant falls ill. *BBC*. <https://www.bbc.com/news/world-54082192>
- Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, 26(2), 198-212.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wu, K. J. (2020, September 10). Safety review underway of AstraZeneca's vaccine trial. *The New York Times*. <https://www.nytimes.com/2020/09/10/health/covid-astrazeneca-vaccine-trans.html>
- Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2008). Improving understanding of adjuvant therapy options by using simpler risk graphics. *Cancer*, 113(12), 3382-3390.