

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	N/A
Data analysis	Data analysis was conducted in R version 4.3.2. Packages used: bayesplay (version 0.9.3), data.table (version 1.16.4), emmeans (version 1.10.6), formattable (version 0.2.1), ggmagnify (version 0.4.1.9000), ggplot2 (version 3.5.2), ggsignif (version 0.6.4), ggtext (version 0.1.2), jtools (version 2.3.0), kableExtra (version 1.4.0), lubridate (version 1.9.4), ltm (version 1.2-0), nFactors (version 2.4.1.2), plotly (version 4.10.4), plyr (version 1.8.9), pwr (version 1.3-0), randomizr (version 1.0.0), sjPlot (version 2.8.17), stringr (version 1.5.1), table1 (version 1.4.3), tidyverse (version 2.0.0) Code availability: Code necessary to reproduce all results reported here are available upon publication on OSF at https://osf.io/mqbx/.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Patient data were extracted from Geisinger Electronic Health Records. Survey data were collected via Qualtrics.

Data availability: De-identified data necessary to reproduce all results reported here are posted on OSF at <https://osf.io/mqbux/>. Data are openly available indefinitely for anyone who wishes to access them for any purpose. All data are posted at the individual level, except vaccination timing data (i.e., vaccination dates) which are aggregated by experimental condition because dates of service are Protected Health Information under 45 CFR §164.514(b). Related documents available include study protocols and statistical analysis plans (available on ClinicalTrials.gov registrations, Study 1: <https://clinicaltrials.gov/ct2/show/NCT04323137>; Study 2: <https://clinicaltrials.gov/ct2/show/NCT05009251>; Study 3: <https://clinicaltrials.gov/ct2/show/NCT05509283>; Study 4: <https://clinicaltrials.gov/ct2/show/NCT05509270>), and analytic code (available with the data at <https://osf.io/mqbux/>).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Field studies: Patient sex for all field studies was extracted from the Geisinger electronic health record. Regressions reported in the main text do not include demographic covariates. However, we reported the primary outcome as a function of sex for all studies in Supplementary Information (SI) section 8. We also ran OLS regressions for each study to test whether intervention effects differed as a function of sex. Interaction effects were not significant for any of the studies, suggesting that our interventions were similarly effective in males and females.

Online pre-tests and post-test: Participants self-reported their sex on the survey by responding to the question "What is your sex?" We report participant sex for each pre-test and post-test in SI sections 9 and 10.

Reporting on race, ethnicity, or other socially relevant groupings

Field studies: Patient race and ethnicity for all field studies were extracted from the Geisinger electronic health record. Regressions reported in the main text do not include demographic covariates. However, we reported the primary outcome as a function of patient race and ethnicity for all studies in SI section 9. To test whether intervention effects varied as a function of race/ethnicity, we created a binary joint race/ethnicity variable, where patients who were both white and non-Hispanic were coded as 1 and those who were not both white and non-Hispanic, or for whom race or ethnicity was unknown, were coded as 0. In Studies 1–3, there were no significant interactions between study arm and this race/ethnicity variable. In Study 4, the race/ethnicity variable significantly interacted with intervention arm, such that patient portal messages and letters, when compared with no messages, were differentially associated with significantly more vaccinations among white non-Hispanic patients.

Online pre-tests and post-test: Participants self-reported their race/ethnicity, in response to the question "What race/ethnicity do you consider yourself?" We report participant race/ethnicity for each pre-test and post-test in SI sections 9 and 10.

Population characteristics

See below for a summary of population characteristics. See the SI for reporting on covariates (age, sex, race, ethnicity) as they interacted with outcomes of interest.

Recruitment

Participants were automatically enrolled in the field studies if they met inclusion criteria listed in SI section 1. There was no potential for self-selection bias, because participants did not have the opportunity to opt into or out of the study. However, in order to meet our inclusion criteria, participants needed to be considered active Geisinger Patients at the time of the study (that is, they previously attended at least one Geisinger primary care appointment, and had at least one additional encounter with Geisinger in the previous two years). The results may not generalize to high-risk patients who are not similarly active at a health system. Additionally, Geisinger is a large and primarily rural health system; results may not generalize to patients in urban settings.

Online pre-tests and post-test: participants were recruited on Amazon MTurk or Prolific and therefore self-selected into those studies. Therefore, the results of the pre-tests or post-test may not generalize to the broader population or to the samples randomized for the main studies.

Ethics oversight

The Geisinger Institutional Review Board (IRB) approved Studies 1 and 2 and waived the requirement for consent. The Geisinger IRB determined that Studies 3 and 4 were health care operations projects and did not constitute human subjects research, and a follow-up research analysis was exempt from review or the requirement for informed consent under 45 CFR §46.104(d)(4)(iii).

The Geisinger IRB also approved the online pre-tests and post-test.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	All RCTs and surveys use experimental designs, with randomization to parallel arms. The data are quantitative.
Research sample	Participants in the RCTs were patients at Geisinger, a healthcare system with hospitals and clinics spanning a large section of Pennsylvania, USA. Mean ages ranged from 53 to 57 across studies, patients were predominantly female (62-67%), white (94-95%) and non-Hispanic (94-97%); full demographic information specific to each study is provided in SI section 3. Patients were selected to be at high risk for flu and flu-related complications among patients at Geisinger. The sample is not nationally representative, but data may be generalizable to other large, rural healthcare settings trying to encourage vaccination in patients at high risk for flu and flu-related complications. The surveys were run on Amazon MTurk or Prolific. Pre-tests 1a, 2 and the post-test targeted participants age 50 and older in hopes of testing material on participants with a similar mean age to patients in our field studies. Mean ages for these studies ranged from 57-58, and participants were predominantly female (58-61%) and white (78-84%). Pre-test 1b did not target older participants (minimum age 18) in hopes of quickly enrolling a larger sample to test many versions of risk message framing. The mean age was 37, and the sample was 48% female and 68% white. Full demographic information for each individual survey is available in SI sections 9 and 10. Survey samples are not representative, though the samples for Pre-tests 1a, 2 and the post-test are similar in age and sex to the field trials.
Sampling strategy	In our field studies, enrollment was fixed and limited to the number of patients determined by the predictive model to be within high ranges of risk considered appropriate for intervention. Minimum Detectable Effects were calculated by considering the anticipated sample size (accounting for available high-risk patients) and the minimal number of arms needed to answer primary questions of interest while detecting the practically achievable effects anticipated for low-touch nudge messages such as ours. Full power analyses for all field trials can be found in the Methods section of the manuscript. For pre-test surveys, no sample size calculation was performed; sample size was determined based on prior similar surveys from our group. For the post-test survey, we included 700 participants per condition (high risk only, medical record review, computer algorithm analysis), which allowed for a 3% margin of error (95% confidence intervals). Full power analysis information can be found on the preregistration (https://osf.io/q69pj).
Data collection	Field trial messages were deployed and implemented by Geisinger using systems that are used in the delivery of normal health system messages. Researchers were not blind to experimental conditions, but there was no direct contact between the research team and patients in the study. Data were recorded in the Geisinger electronic health system through normal interactions with the health system (e.g., flu shots were documented during typical appointments). Subsequently, data for all field trials were extracted from the Geisinger electronic health record via SQL data pulls. Survey data were collected on Qualtrics through Amazon MTurk or Prolific.
Timing	Study 1: September 21, 2020 – May 31, 2021 Study 2: September 9, 2021 – November 5, 2021 Study 3: September 13, 2022 – October 26, 2022 Study 4: September 6, 2022 – October 4, 2022 Pre-test 1a: August 14, 2020 – August 15, 2020 Pre-test 1b: August 14, 2020 Pre-test 2: August 3, 2021 – August 14, 2021 Post-test: April 4, 2024 – April 6, 2024
Data exclusions	Study 1: 589 patients were excluded for being contraindicated for flu vaccine (part of the original exclusion criteria; these patients should not have been enrolled); 7,415 patients were excluded for being vaccinated prior to the study start date (preregistered exclusion) Study 2: 563 patients were excluded for being contraindicated for flu vaccine (part of the original exclusion criteria; these patients should not have been enrolled); 1,096 patients were excluded for being vaccinated prior to the study start date (preregistered exclusion; these patients were included in the preregistered analysis reported in section 8.2.2); 105 patients were excluded from analysis in the main text because they died during the study (preregistered exclusion; these patients were included in the preregistered analysis reported in section 8.2.2); 816 patients were excluded from analysis in the main text because they were sent messages including protective risk reasons, or would have been sent messages with protective risk reasons if they had been randomized to an arm that included reasons (preregistered exclusion; these patients were included in the preregistered analysis reported in section 8.2.2) Study 3: 1,542 patients were excluded for being vaccinated prior to the study start date (preregistered exclusion) Study 4: 516 patients were excluded for being vaccinated prior to the study start date (preregistered exclusion)

CONSORT diagrams in SI section 2 show the above exclusions separately by study arm

Pre-test 1a: 5 participants were excluded for failing an attention check; 4 who were excluded for finishing the survey in under 2 minutes; and 3 were excluded for not completing the survey

Pre-test 1b: 3 participants were excluded because they were sent an old survey version with a typo; 37 participants were excluded for failing an attention check; 49 participants were excluded for finishing the survey in under 2 minutes; 14 indicated they did not want us to use their responses in data analysis; and 23 were excluded for not completing the survey

Pre-test 2: 8 participants were excluded for failing an attention check; 2 were excluded for finishing the survey in under 2 minutes; 33 were excluded for not completing the survey; and 5 indicated they did not want us to use their responses in data analysis; and 23 were excluded for not completing the survey

Pre-tests were not preregistered, so none of the exclusions were pre-defined.

Post-test: 53 participants were excluded from our primary analysis for responding incorrectly to a multiple choice attention check question on the page following the question of interest ("If you are paying attention to this question, select 'Moderately surprising'; N = 2 incorrect responses, N = 51 no response). These exclusions were preregistered (<https://osf.io/q69pj>)

Non-participation

No participants dropped out of the study

Randomization

Field studies: Randomization was performed in R

Online pre-tests and post-test: Randomization was performed using the Qualtrics randomizer

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

- n/a Involved in the study
- ☒ ☐ Antibodies
 - ☒ ☐ Eukaryotic cell lines
 - ☒ ☐ Palaeontology and archaeology
 - ☒ ☐ Animals and other organisms
 - ☐ ☒ Clinical data
 - ☒ ☐ Dual use research of concern
 - ☒ ☐ Plants

- n/a Involved in the study
- ☒ ☐ ChIP-seq
 - ☒ ☐ Flow cytometry
 - ☒ ☐ MRI-based neuroimaging

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

Study 1: <https://clinicaltrials.gov/ct2/show/NCT04323137>; Study 2: <https://clinicaltrials.gov/ct2/show/NCT05009251>; Study 3: <https://clinicaltrials.gov/ct2/show/NCT05509283>; Study 4: <https://clinicaltrials.gov/ct2/show/NCT05509270>

Study protocol

Protocols for all RCTs are posted on clinicaltrials.gov at the links above.

Data collection

Data from the clinical field studies were conducted by pulling data from the Geisinger electronic health record. Flu shots were extracted for the following dates:

Study 1: September 21, 2020 – May 31, 2021

Study 2: September 9, 2021 – November 5, 2021

Study 3: September 13, 2022 – October 26, 2022

Study 4: September 6, 2022 – October 4, 2022

Outcomes

The primary outcome for all four field studies was flu vaccination, which was preregistered on clinicaltrials.gov for all studies. This outcome was assessed by pulling all flu shot records from the Geisinger electronic health record for patients in each study in the preregistered time frame (see below). The present manuscript reports only on this primary outcome. Preregistered secondary and exploratory outcomes fall outside the scope of the current manuscript. Clinical leaders at Geisinger assisted the study team in determining the outcomes of interest for all studies and outcomes.

The preregistered primary outcome time frames for flu vaccination varied across studies (Study 1: September 21, 2020 – May 31, 2021; Study 2: September 9, 2021 – November 5, 2021; Study 3: September 13, 2022 – October 26, 2022; Study 4: September 6, 2022 – October 4, 2022). Time frames were shortened after Study 1 because the team observed that the effect of messages was largest in the weeks immediately following the messages. Additionally, precise time frames varied across Studies 2–4 due to practical considerations that differed across studies. For Study 2, the primary outcome time frame was 2 weeks after the final SMS send date.

In Study 3, because print letters were inadvertently sent on 2 separate consecutive days, the primary outcome time frame was within 6 weeks of the patient's study start date, so all patients could be monitored for the same length of time (patients in the passive control condition were randomly assigned to one of the two study start dates). For Study 4, because all messages were sent within two days, we shortened the time frame to within 4 weeks of the study start date.