Communicating overall and statistical uncertainty: A randomized trial of ways to report the results of a study on wearing glasses to reduce COVID risk

steven woloshin, Christine Holst, Andrew Oxman, Christopher Rose, Sarah Rosenbaum, Heather Menzies Munthe-Kaas

Submitted to: JMIR Public Health and Surveillance
on: June 04, 2024

Table of Contents

Original Manuscript **5**
Supplementary Files **48**
Figures **49**
Figure 1 ..... **51**
Figure 2 ..... **53**
Figure 3 ..... **54**
Figure 4 ..... **55**
Figure 5 ..... **56**
Figure 6 ..... **57**
Figure 7 ..... **58**
Multimedia Appendixes **59**
Multimedia Appendix 1 ..... **60**

Communicating overall and statistical uncertainty: A randomized trial of ways to report the results of a study on wearing glasses to reduce COVID risk

steven woloshin[1, 2] MD, MS; Christine Holst[2, 3] MPhil; Andrew Oxman[2, 3] MD; Christopher Rose[1, 3] PhD; Sarah Rosenbaum[2, 3] PhD; Heather Menzies Munthe-Kaas[2, 3] PhD

[1]Dartmouth Institute for Health Policy and Clinical Practice Geisel Medical School west lebanon US
[2]Lisa Schwartz Foundation for Truth in Medicine Norwich US
[3]Centre for Epidemic Interventions Research Norwegian Institute of Public Health Oslo NO

Corresponding Author:
steven woloshin MD, MS
Dartmouth Institute for Health Policy and Clinical Practice
Geisel Medical School
1 medical center drive
west lebanon
US

Abstract

Background: Communicating uncertainties to the public about the results of health research is challenging.

Objective: To test alternative ways to express overall and statistical uncertainty, using a study about glasses to reduce COVID risk as an example.

Methods: Design Two online, parallel, individually randomized trials (both 3×2 factorial designs), randomized participants to different ways of presenting overall uncertainty ("GRADE language", "plain language", or "no explicit language") and statistical uncertainty (margin of error or none).
Participants Non-glasses wearing adults, recruited from web-based research panels.
Main outcome Understanding overall and statistical certainty of evidence.
Secondary outcomes Perceived benefit, intended glasses-wearing behavior, assessment of provided information.
Analysis Trials analyzed separately and combined in meta-analysis.

Results: In the U.S. and Norwegian trials, 730 and 497 individuals were randomized; data for 543 and 452 analyzed (i.e., after exclusions, mostly failing attention checks): overall mean age 41 years (SD=15.4), 49% female, 48% with at least a college degree.
Without margin of error, plain vs. no explicit language increased correct understanding of how sure one can be about the effect of wearing glasses: [US] 20.9% to 37.4% and [Norway] 42.5% to 52.6% (combined odds of correct understanding, 1.82, (95% CI 1.16 to 2.88)). Effect of GRADE vs. no explicit language was less certain: OR 1.29 (95% CI 0.80 to 2.06).
Adding margin of error reduced understanding of plain (OR 0.51, 95% CI 0.27 to 0.99) and GRADE language (OR 0.82, 95% CI 0.42 to 0.1.60), with a shift from perceiving the evidence as "mixed but more unsure than sure" (correct) to "very unsure".
Adding margin of error increased the proportion correctly summarizing the glasses effect (i.e., "wearing glasses may reduce the chance of getting Covid a little, but might increase it a little") from 1%-3% correct to 21%-31% for the 3 kinds of language (overall OR 10.84, 95% CI 3.74 to 31.44).
Plain vs. no explicit language increased the proportion correctly agreeing evidence was insufficient to be sure about the effect of glasses in the U.S. trial: 67.4% to 88.9%, risk difference 21.4%, 95% CI 9.8% to 33.1% but not in the Norwegian trial: 81.2% to 81.6 %, risk difference 0.3%, 95% CI -11.9% to 12.5%. The effect of GRADE vs. no explicit language was less certain (67.4% to 79.7% [US], and 81.2% to 79.5% [Norway]).

Conclusions: Plain but not GRADE language was better than no explicit language in helping people understand overall certainty of evidence, though most participants did not correctly understand how sure they could be. Reporting margin of error reduced understanding of overall uncertainty by making people feel the evidence was even less certain. Reporting margin of error improved interpretation of statistical uncertainty around the effect of glasses, but only for a minority of participants. Clinical

Trial: ClinicalTrials.gov. NCT05642754

(JMIR Preprints 04/06/2024:62828)
DOI: https://doi.org/10.2196/preprints.62828

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Article type: Randomized Trial

Communicating overall and statistical uncertainty: A randomized trial of ways to report the results of a study on wearing glasses to reduce COVID risk

Steven Woloshin[1,2], Christine Holst[2,3], Andrew D. Oxman[2,3], Christopher J. Rose[2,3], Sarah E. Rosenbaum[2,3], Heather Munthe-Kaas[2,3]

[1]Center for Medicine in the Media, Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, New Hampshire, United States
[2]The Lisa Schwartz Foundation for Truth in Medicine, Norwich, Vermont, United States
[3]Centre for Epidemic Interventions Research, Norwegian Institute of Public Health, Oslo, Norway

Corresponding author:
Steven Woloshin, MD, Professor of Medicine, Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH, USA
E: steven.woloshin@dartmouth.edu

Trial registration: ClinicalTrials.gov. NCT05642754

Protocol posted on Zenodo. https://doi.org/10.5281/zenodo.7428981

Word count = 5321

3 Tables
5 Figures

ABSTRACT. (word count = 450)

Background Communicating uncertainties to the public about the results of health research is challenging.

Objective To test alternative ways to express overall and statistical uncertainty, using a study about glasses to reduce COVID risk as an example.

Design Two online, parallel, individually randomized trials (both 3×2 factorial designs), randomized participants to different ways of presenting *overall uncertainty* ("GRADE language", "plain language", or "no explicit language") and *statistical uncertainty* (margin of error or none).

Participants Non-glasses wearing adults, recruited from web-based research panels.

Main outcome Understanding overall and statistical certainty of evidence.

Secondary outcomes Perceived benefit, intended glasses-wearing behavior, assessment of provided information.

Analysis Trials analyzed separately and combined in meta-analysis.

Results In the U.S. and Norwegian trials, 730 and 497 individuals were randomized; data for 543 and 452 analyzed (i.e., after exclusions, mostly failing attention checks): overall mean age 41 years (SD=15.4), 49% female, 48% with at least a college degree.

Without margin of error, plain vs. no explicit language increased correct understanding of how sure one can be about the effect of wearing glasses: [US] 20.9% to 37.4% and [Norway] 42.5% to 52.6% (combined odds of correct understanding, 1.82, (95% CI 1.16 to 2.88)). Effect of GRADE vs. no explicit language was less certain: OR 1.29 (95% CI 0.80 to 2.06).

Adding margin of error reduced understanding of plain (OR 0.51, 95% CI 0.27 to 0.99) and GRADE language (OR 0.82, 95% CI 0.42 to 0.1.60), with a shift from perceiving the evidence as “mixed but more unsure than sure” (correct) to “very unsure”.

Adding margin of error increased the proportion correctly summarizing the glasses effect (i.e., "wearing glasses may reduce the chance of getting Covid a little, but might increase it a little") from 1%-3% correct to 21%-31% for the 3 kinds of language (overall OR 10.84, 95% CI 3.74 to 31.44).

Plain vs. no explicit language increased the proportion correctly agreeing evidence was insufficient to be sure about the effect of glasses in the U.S. trial: 67.4% to 88.9%, risk difference 21.4%, 95% CI 9.8% to 33.1% but not in the Norwegian trial: 81.2% to 81.6 %, risk difference 0.3%, 95% CI -11.9% to 12.5%. The effect of GRADE vs. no explicit language was less certain (67.4% to 79.7% [US], and

81.2% to 79.5% [Norway]).

Conclusions Plain but not GRADE language was better than no explicit language in helping people understand overall certainty of evidence, though most participants did not correctly understand how sure they could be. Reporting margin of error reduced understanding of overall uncertainty by making people feel the evidence was even less certain. Reporting margin of error improved interpretation of statistical uncertainty around the effect of glasses, but only for a minority of participants.

Key Words

Communication, GRADE language, Statistical uncertainty, Overall certainty

Introduction

Public health messaging matters – it shapes how people understand important health risks and what can be done to mitigate them. Ideally such communications are evidence-based – but even the best communications can fail if messaging is executed poorly, as has been seen during the COVID epidemic. [1]

Decision scientists have articulated basic principles for effective health communication, for example using simple and familiar wording, clear visual design, presenting structured comparisons of alternatives, and careful testing in the target audiences. [2-5]

A recent randomized trial assessed the effect of these principles on communication effectiveness in the context of COVID-19 home test kit instructions, using a real example. [6] The trial showed that individuals randomized to instructions that did not follow best decision science principles (the actual FDA-authorized instructions) vs. those that did (carefully pre-tested intervention instructions) were more likely to fail to quarantine appropriately (33% vs. 14% failed; 95%CI for the 19% difference, 6%to 31%).

Evidence from randomized trials documents the importance of both the format and content of health messaging. Formatting examples include how use of percentages (e.g., 10%) vs. frequency formats (e.g., 10 in 100) can improve comprehension, [7] how absolute vs. relative risk measures for communicating treatment effects are better understood [3,8] and help people make decisions more consistent with their values. [9]

Content examples include the importance of presenting both benefits and harms when describing interventions, [10] and highlighting study limitations such as how simple nondirective explanations about surrogate outcomes and newly approved drugs enhance evidence-based decision making about prescription drugs. [11]

There is also evidence supporting the importance of communication about the uncertainty of research findings in both the professional literature, [12-15] and in plain language summaries for the public. [16] This includes both *statistical uncertainty* (i.e., imprecision)[15] and the *overall uncertainty*, due to the risk of bias, inconsistency, indirectness, and publication bias. [12,14] However there are still open questions about the best formats and language for presenting both kinds of uncertainty, and how people understand and react to such information.

A recent pair of trials [17] found that including “quality cues” in communications tempered the publics’ tendency to assume that quality of evidence presented without such cues is high (when it was not), and reporting that evidence quality was low decreased trust, perception of intervention efficacy, and likelihood of adopting it.

The certainty of the evidence can affect the decisions that people make. If the purpose of a message[18] is to inform people rather than to persuade them, [19] it is necessary to include information about the certainty of the evidence. Not doing so can be misleading.

The present study is designed as a proof-of-concept exercise to develop and test a communication meant to summarize the results of a randomized trial, specifically a study examining the effects of wearing glasses on the chance of developing COVID-19. [18] This work is part of a larger project to establish a "message lab" that will promote best practices in message development (i.e., attending to the foregoing communication principles and evidence), facilitate user testing and conduct randomized trials assessing the effects of public health messages on the publics’ understanding of the messages, beliefs, decisions, and behaviours.

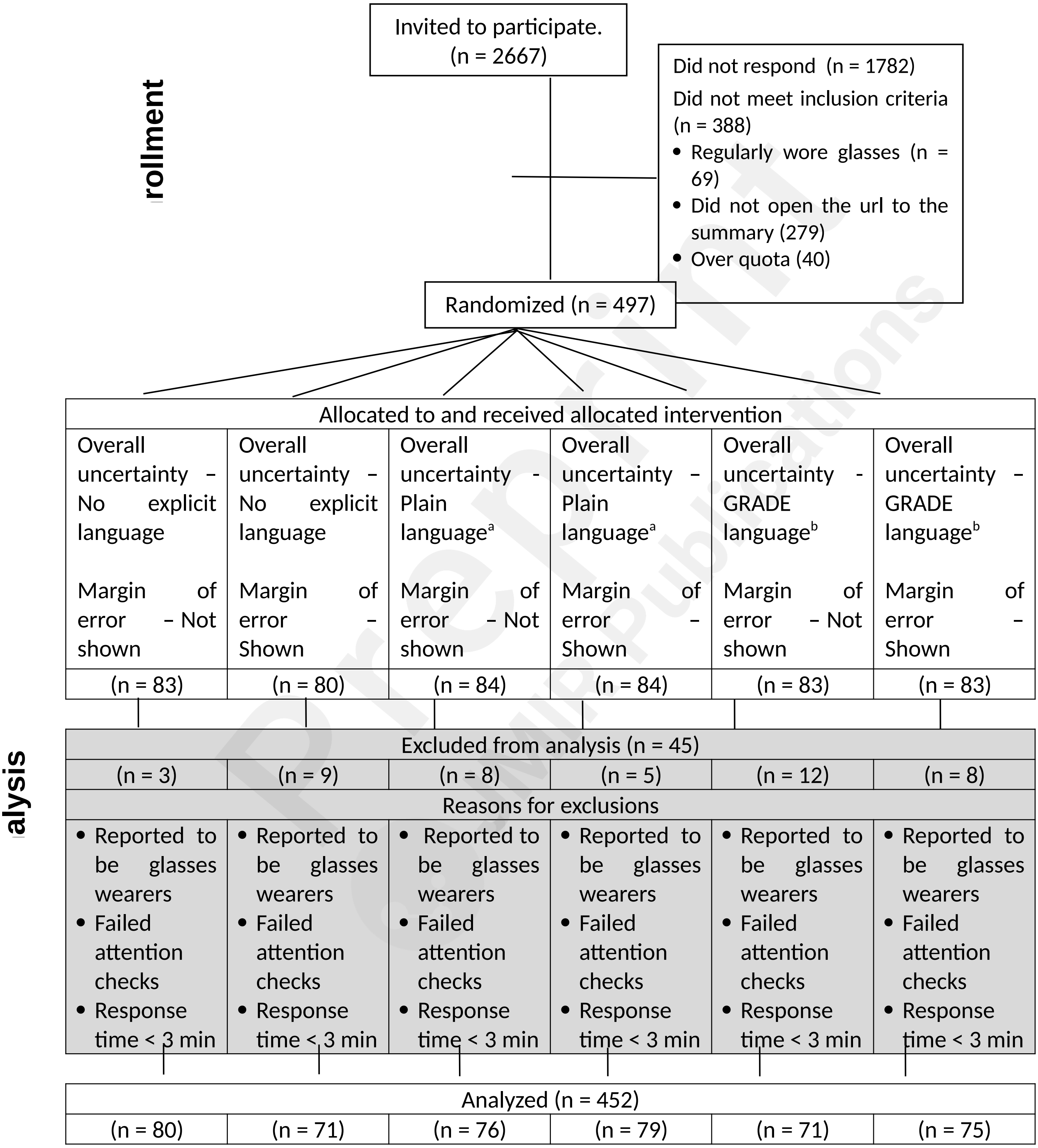
Methods

Design

We designed an online, parallel group, individually randomized, pragmatic trial to compare the effects of different ways of communicating uncertainty when reporting the results of a randomized trial to the public. We used a published trial assessing the effect of wearing glasses on the risk of being infected with Covid as our example. Since we were interested in 2 kinds of uncertainty (statistical and overall [e.g., risk of bias, inconsistency, indirectness]), we used a 3 x 2 factorial design with 3 kinds of overall uncertainty [GRADE, Plain or No explicit language] with or without reporting the margin of error [i.e., confidence interval] around the Covid trial's main result. This resulted in 6 (3x2) groups differing in how the Covid study was summarized (see Figure 1 for study design, and Figure 2 for the language used in each group).

Figure 1. Flow diagram of study participants in the (a) Norwegian and (b) U.S. trials

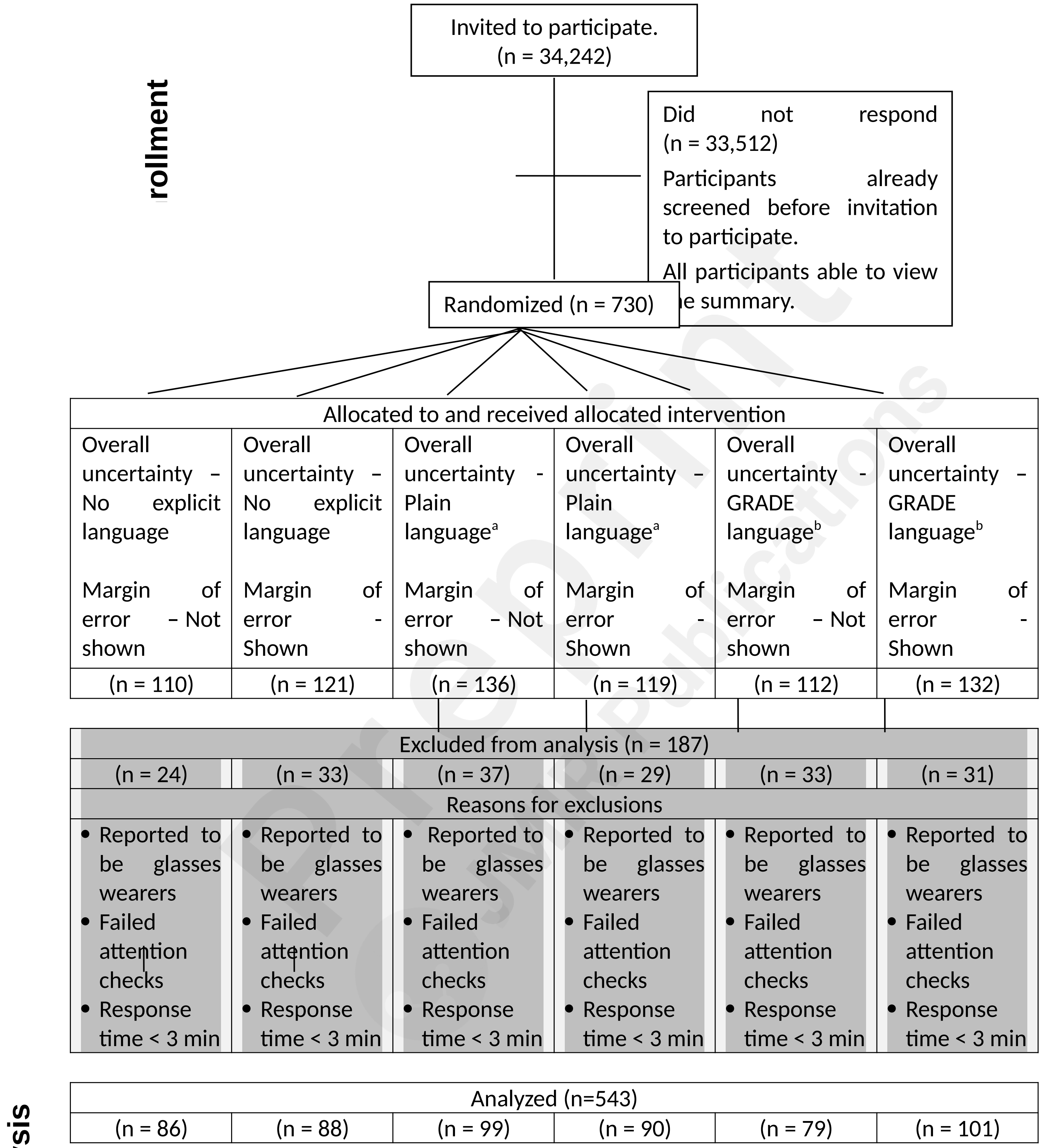
1a. Norwegian trial.



[a]A less formal expression of the overall uncertainty language used in ordinary or familiar conversation, corresponding to the same GRADE assessment of the certainty of the evidence.

[b]Based on the Cochrane Effective Practice and Organisation of Care Group's guidance for communicating the certainty of evidence based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to assessing the certainty of evidence[34]

1b. U.S. trial



[a]A less formal expression of the overall uncertainty language used in ordinary or familiar conversation, corresponding to the same GRADE assessment of the certainty of the evidence.

[b]Based on the Cochrane Effective Practice and Organisation of Care Group's guidance for communicating the certainty of evidence based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to assessing the certainty of evidence[34]

Figure 2. (a) Overall uncertainty language about the possible benefit of glasses to reduce Covid used in each intervention group in the U.S. trial (the Norwegian version is in Appendix). (b) Margin of error language was presented in the same way in the 3 versions reporting it.

(a) Overall uncertainty

	What are the effects of wearing glasses to reduce the chance of getting COVID?
GRADE language*	What is the benefit? Wearing glasses may slightly reduce your chance of getting COVID (⊕⊕◯◯ low certainty evidence). What are the downsides? Wearing glasses probably does not increase your chance of any major harms, such as a serious fall due to reduced vision (⊕⊕⊕◯ moderate certainty evidence). But some people are irritated by foggy glasses, and some feel silly wearing them.
Plain language	What is the benefit? Wearing glasses may reduce your chance of getting COVID a little – but we are not very confident about this. What are the downsides? We are somewhat confident that wearing glasses does not cause important harms such as a serious fall due to reduced vision. But some people are irritated by foggy glasses, and some feel silly wearing them.
No explicit language	What is the benefit? Wearing glasses may reduce your chance of getting COVID a little. What are the downsides? Wearing glasses probably does not cause important harms such as a serious fall due to reduced vision. But some people are irritated by foggy glasses, and some feel silly wearing them.

(b) *Margin of error*

	Here's what happened:
Reported	Over the next 2 weeks, slightly fewer people told to wear glasses tested positive for COVID: 9.6% compared to 11.5% of those told not to wear glasses. That's a difference of 1.9%. Accounting for the play of chance (i.e., the margin of error), glasses might *reduce* the risk of COVID by as much as 3.9% but might *increase* it by as much as 0.1%.
Not reported	Over the next 2 weeks, slightly fewer people told to wear glasses tested positive for COVID: 9.6% compared to 11.5% of those told not to wear glasses. That's a difference of 1.9%.

Recruitment of participants

We implemented the design in 2 nearly identical trials in Norway and the U.S. (differences noted below). For each trial we used quota sampling: from a web-based platform of volunteer research subjects in the U.S. Trial [20] and from an independent commercial research agency [21] with a panel of 120,000 people living in Norway. Participants were eligible if they were age 18 and older, said they did not regularly wear glasses (based on a pre-screen feature in Prolific and a screener question in Opinion). Participants were literate in English (used in the U.S. version) or Norwegian (used in the Norwegian version). Both platforms apply processes to prevent bots and fraudulent participants [22] and to further increase data quality we built in attention and comprehension checks. [23] Participants were paid $8.40 and $4.30 in the Norwegian and U.S. trials respectively.

Randomization

Eligible participants were randomized in equal numbers to one of the six groups above. In the US, participants were randomized from the questionnaire, after clicking on the link to the summary. In Norway, the participants were randomized to each summary by the research agency.

The participants were then shown one of six Covid trial summaries (in English or Norwegian depending on the trial) according to their randomized group (Figure 1).

Intervention

The language used for the six Covid trial summary versions are shown in Figure 2 and the full summaries are in Appendix 1. Each differed in how overall uncertainty about the benefits and harms were presented (i.e., GRADE, plain or no explicit language) and whether the margin of error was reported.

The summaries were drafted in English and translated to Norwegian, by two researchers. Note that the summaries underwent multiple rounds of user testing and modifications as needed prior to the trial. We used human-centered design methods to develop the text to be presented to participants in the trial. [24] Initially, we invited four native English speaker and three native Norwegian speakers to unmoderated user testing using Loop11, a digital user experience platform. [25] The user test participants (five female, two male, two with secondary school education, five with post-secondary education, two living in Canada, five living in Norway, three over 55 years old and four between 30 and 45 years old) were introduced to the project, and asked to imagine themselves in the following scenario: "Imagine you hear that glasses may reduce your chance of becoming infected with Covid. You go online to find out more information and find a website that says…[insert Covid trial summary version 1, 2 or 3]."

They were then asked to read five pieces of text: three versions of the summaries with progressively more information (e.g., the first version was the shortest summary with no text related to certainty or margin of error, the second version had text related to certainty but not margin of error, and the third version had text related to certainty and margin of error), the trial questions, and the text that would be available via hyperlinks and hovertext in the summaries. For each piece of text, the user-test participant was asked for their first impressions and then a series of follow-up questions related to content, font, format, language, and anything else.

We revised the summaries according to feedback. The translations were reviewed by a third researcher. The English summaries were presented to one more participant in English (moderated, in-person user experience interview) using the same format as the first four interviews. The summaries were subsequently revised, and changes were made also to the Norwegian summaries.

We gathered feedback on the Norwegian summaries from colleagues and conducted two in-person user experience interviews with two native Norwegian speakers using the same question guide as the English user experience interviews. Revisions were made according to feedback, and where appropriate these revisions were back translated into the English versions of the Covid trial summaries.

We also gathered feedback from user experience participants on other materials related to the process of participating in the randomized trial (e.g., from invitation to participate, to the text sent after participants were finished the questionnaire).

Outcomes

We defined three binary co-primary outcomes to measure understanding of overall uncertainty and one to measure understanding of statistical uncertainty (precision of the effect estimate for the benefit). Each of the three outcomes for overall uncertainty will be measured by comparing participants' answers to questions about uncertainty to expected (correct) answers, based on the size of the effects and the certainty of the evidence (exact question language and "correct" answers shown in results tables and figures, the full questionnaire is in the Appendix 2).

Three *co-primary outcomes* are as follows (all used 4-point ordinal response sets):

- *Understanding of the uncertainty of the benefit* ("How sure are you about the effect of wearing glasses on your chance of getting COVID?)
- *Understanding of the sufficiency of the evidence* ("Not enough is known to be sure about the effects of wearing glasses to reduce the chance of getting COVID")
- *Understanding of the certainty of important harms* ("Wearing glasses to reduce the chance of getting COVID does not cause important harms")

An additional co-primary outcome was included for the 3 versions reporting the margin of error assessing understanding of statistical uncertainty (i.e., precision of the effect estimates: choose which from among 4 statements was most consistent with the information provided, e.g., wearing glasses may reduce the chance of COVID a little but might reduce it a lot").

Secondary Outcomes included questions about:

- the perceived benefit and harm of wearing glasses to reduce the chance of COVID infections,
- intended behaviour (whether participants would wear glasses to reduce the chance of getting COVID in areas with high and low COVID infection rates),
- perceptions of the trustworthiness, adequacy, clarity, and helpfulness of the information, and likelihood of sharing it with others
- decisional conflict [26]

Statistical analysis.

The U.S. and Norwegian trials were analyzed in the same way, except as noted. We excluded participants who completed the survey in less than three minutes, reported to be glasses wearers or failed the attention checks. Data were duplicated for some U.S. participants who appeared to have submitted the same responses multiple times in a short period. We assumed these participants clicked the submit button several times or refreshed their browser, so analyzed only the first data submitted by these participants. All analyses were performed before unblinding as prespecified according to the intention-to-treat principle - all randomized participants meeting the inclusion criteria were included and analyzed in the arms to which they were randomized (the trial protocol was posted, prior to recruitment[27]).

All outcomes were binomial. We used logistic regression to estimate odds ratios for the two treatments and the interactions between them. Model fit was assessed using the Hosmer-Lemeshow test. To aid interpretation, we re-expressed odds ratios as risk differences, accounting for statistical uncertainty on baseline odds, main effects, and interactions. No data were missing for any of the participants meeting the inclusion criteria.

We used fixed effects meta-analysis to pool estimates across the trials and obtain overall estimates of effect and assess country as a potential effect modifier. We performed prespecified subgroup analyses for both trials to explore differences in treatment effect with respect to numeracy and, in the U.S. trial, saliency. Numeracy was defined as scoring 3 (versus less than 3) on a validated instrument. [28] Saliency was defined as being very or extremely worried about getting COVID and it being very or extremely important to take actions to reduce the chance of getting COVID. It was not possible to perform the analysis for saliency for the Norwegian trial because only 20 participants met the saliency criteria.

We also performed non-prespecified analyses for potential effect modification. In the U.S. trial we explored effect modification by response time (responding within seven minutes versus more than seven minutes). The choice of seven minutes was data-driven, chosen to be a whole number close to the median response time. We also explored effect modification by education level (having graduated versus not having graduated from university).

We present two-sided 95% confidence intervals and p-values, where applicable, throughout. Meta- and subgroup analyses are presented using forest plots, with p-values testing null hypotheses of homogeneity (no difference between estimates for the two trials or no effect modification). All analyses were performed using Stata 18 (StataCorp LLC, College Station, Texas, USA).

Ethical considerations

This study was considered for ethical approval by the Regional Committee for Medical and Health Research Ethics and was found not to require ethical approval since it falls outside the Committees mandate under the Health Research Act (reference 557972).

The study was also deemed “Exempt” from further review by the Committee for the Protection of Human Subjects (CPHS, STUDY00032615), the Institutional Review Board (IRB) at Dartmouth

College.

Patient and Public Involvement

Patients were involved in user testing during the development of the survey and COVID trial summaries.

Competing interests

None declared.

Funding

This research is funded by the Norwegian Institute of Public Health.

Results

We invited 2667 Norwegians to participate in the trial, 1782 did not respond, 388 were not eligible, and 497 were randomized to one of the six intervention groups (Figure 1a). A total of 45 participants were excluded from the analysis for the reasons shown in Figure 1a.

We invited 34,242 people in the U.S. to participate in the trial, 33,512 did not respond and 730 were randomized to one of the six intervention groups (Figure 1b). A total of 187 participants were later excluded from the analysis for the reasons shown in Figure 1b.

The average age of participants was 45 in the Norwegian trial and 38 in the U.S. trial (Table 1).

Table 1. Characteristics of the participants

	Norwegian trial n=452	U.S. trial n=543	Total n=995
Age			
mean (sd) [years]	45.0 (16.8)	38.0 (13.2)	41.2 (15.4)
18-29 years	24%	29%	27%
30-39 years	16%	36%	27%
40-49 years	18%	18%	18%
50-59 years	19%	5%	12%
60-69 years	13%	11%	12%
70+ years	9%	2%	5%
Sex			
Female	50%	48%	49%
Male	50%	52%	51%
Race/Ethnicity			
Asian	n/a	5%	
Black	n/a	8%	
White	n/a	79%	
Other	n/a	4%	
Multiple	n/a	4%	
Employment status			
Full time	46%	39%	
Part time	8%	12%	
Unemployed/seeking work	n/a	9%	
Missing	5%	24%	
Unpaid[a]	n/a	10%	
Welfare[b]	25%	n/a	
Self-employed	5%	n/a	
Student	11%	n/a	
Other	1%	6%	
Education			
<High school	7%	5%	6%
High school degree	29%	17%	22%
Some college	20%	26%	23%
College degree	27%	38%	33%
Graduate or professional school	16%	15%	15%
Numeracy			
<3	52%	52%	52%
3	48%	48%	48%

[a]Includes participants who are homemakers, retired, or disabled
[b]Includes participants in any welfare program; maternity leave, pensioner, un employment benefits

Half of the Norwegians and 48% of the U.S. participants were female. The educational level was higher in the U.S. participants (53% had at least a college degree) than in the Norwegian participants (43% had at least a college degree). About half (52%) of participants in both trials failed to answer all three numeracy questions correctly (see Table 1).

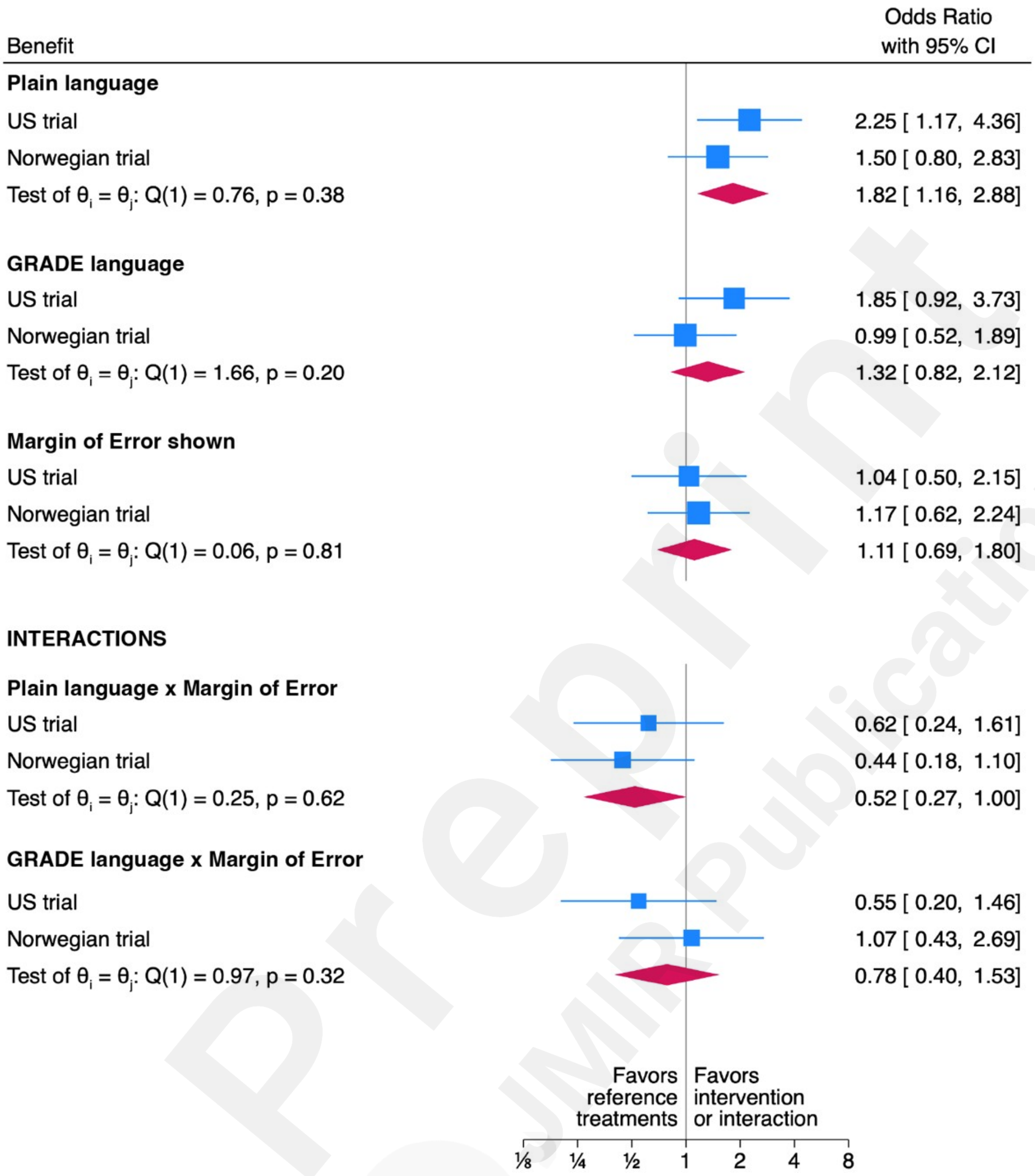
There were only minor differences among the comparison groups in both the Norwegian and U.S. trials (Appendix 3, Tables 3a and 3b).

Understanding of the certainty of the benefit

Overall

Overall, plain (but not GRADE) language compared to no explicit language improved correct understanding of the overall uncertainty of the effect of wearing glasses on the chance of getting COVID (i.e., mixed but more unsure than sure, OR 1.82, 95% CI 1.16 to 2.88) (Figure 3).

Figure 3. Understanding of the certainty of the benefit of wearing glasses*



NOTE: Odds ratios for answering "Mixed but more unsure than sure" to the question "How sure are you about the effect of wearing glasses on your chance of getting COVID?".

When the margin of error was shown together with plain language the effect on understanding of the overall uncertainty was substantially reduced (OR 0.52, 95% CI 0.27 to 1.00) (Figure 3). The reason for this is that participants were more likely to perceive the evidence as “very unsure” rather than “mixed but more unsure than sure” (the correct response) when the margin of error was shown (Appendix 4).

Individual trials

Overall, on average, across all six comparison groups, there were more Norwegian participants who responded that the evidence was very unsure (25.2%) compared to U.S. participants (16.4%) (Appendix 4, Tables 4a and 4b). One reason for this is that “may reduce” was translated to “kan muligens redusere” (“can possibly reduce”) in Norwegian. [29] This was done because “kan” in Norwegian can mean either may or can.

Based on the results of the U.S. trial, using plain language increased correct understanding of how sure we can be about the effect of wearing glasses on the chance of getting COVID from 20.9% to 37.4% (risk difference 16.4%, 95% CI 3.6 to 29.3) (Table 2a). The risk difference for the Norwegian trial was less certain (risk difference 10.1%, 95% CI -5.5 to 25.7) (Table 2b).

Table 2. Risk differences for the understanding the certainty of the benefit of wearing glasses

2a. Norwegian trial

How sure are you about the effect of wearing glasses on your chance of getting COVID?

Very sure
Mixed but more sure than unsure
Mixed but more unsure than sure (correct)
Very unsure

Uncertainty	Margin of Error	n/N (%)[a]	Odds ratio [95% CI][b]	Risk difference [95% CI][c]	*P-value*
No explicit language	Not shown	34/80 (42.5%)	1	0	NA
No explicit language	Shown	33/71 (46.5%)	1.17 [0.62 to 2.24]	3.98 [-11.89 to 19.85]	.623
Plain language	Not shown	40/76 (52.6%)	1.50 [0.80 to 2.83]	10.13 [-5.47 to 25.73]	.206
Plain language	Shown	29/79 (36.7%)	0.78 [0.42 to 1.48]	-5.79 [-20.97 to 9.39]	.456
GRADE language	Not shown	30/71 (42.3%)	0.99 [0.52 to 1.89]	-0.25 [-16.04 to 15.54]	.976
GRADE language	Shown	36/75 (48.0%)	1.25 [0.66 to 2.35]	5.50 [-10.16 to 21.16]	.492

[a]Number (n) of participants randomized to the intervention (N) answering correctly or as anticipated. [b]Odds ratios include the main and interaction effects. [c]Risk differences account for uncertainty on the baseline odds.

2b. U.S. trial

How sure are you about the effect of wearing glasses on your chance of getting COVID?

Very sure
Mixed but more sure than unsure
Mixed but more unsure than sure (correct)
Very unsure

Overall uncertainty	Margin of Error	n/N (%)[a]	Odds ratio [95% CI][b]	Risk difference [95% CI][c]	*P-value*
No explicit language	Not shown	18/86 (20.9%)	1	0	NA
No explicit language	Shown	19/88 (21.6%)	1.04 [0.50 to 2.15]	0.7% [-11.5 to 12.8]	.915
Plain language	Not shown	37/99 (37.4%)	2.25 [1.17 to 4.36]	16.4% [3.6 to 29.3]	.016
Plain language	Shown	25/90 (27.8%)	1.45 [0.73 to 2.91]	6.8% [-5.8 to 19.5]	.292
GRADE language	Not shown	26/79 (32.9%)	1.85 [0.92 to 3.73]	12.0% [-1.5 to 25.4]	.084
GRADE language	Shown	22/101 (21.8%)	1.05 [0.52 to 2.12]	0.9% [-10.9 to 12.6]	.887

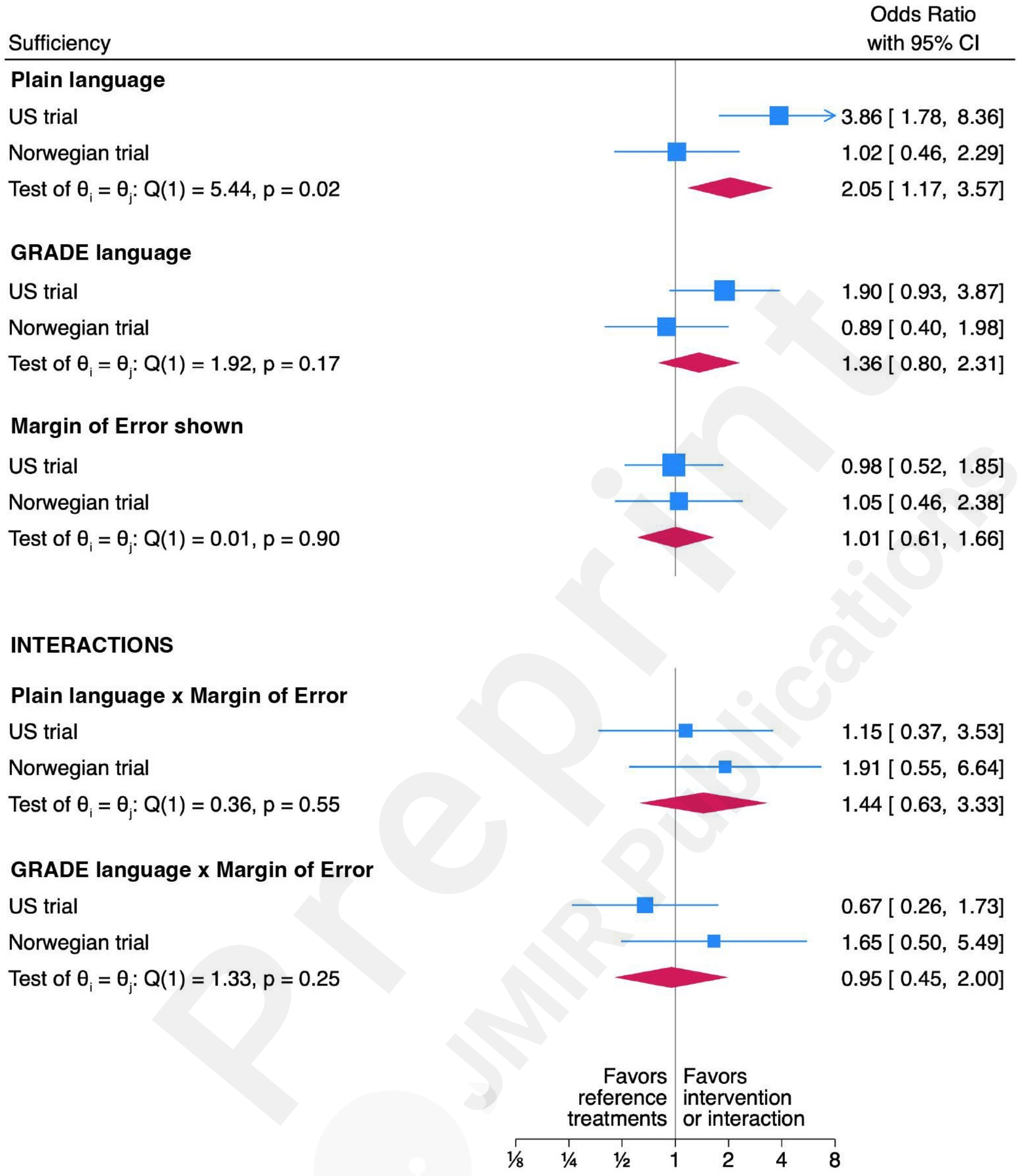
[a]Number (n) of participants randomized to the intervention (N) answering correctly or as anticipated. [b]Odds ratios include the main and interaction effects. [c]Risk differences account for uncertainty on the baseline odds.

Understanding of the sufficiency of the evidence

Overall

Overall, plain compared to no explicit language also improved correct understanding of the sufficiency of the evidence (OR 2.05, 95% CI 1.17 to 3.57) (Figure 4). However, the results of the two trials were heterogeneous (Q(1) = 5.44, p = 0.02). The odds ratio for the Norwegian trial was 1.02 (95% CI 0.46 to 2.29). Whereas the odds ratio for the U.S. trial was 3.86 (95% CI 1.78 to 8.36).

Figure 4. Understanding sufficiency of the evidence*



NOTE: Odds ratios for agreeing or strongly agreeing with "Not enough is known to be sure about the effects of wearing glasses to reduce the chance of getting COVID."

Individual trials

The results of the Norwegian trial for this outcome likely reflect a ceiling effect. 81.2% of participants who were shown no explicit language for overall uncertainty and no margin of error (Table 3a) agreed that not enough was known to be sure about the effects of wearing glasses to reduce the chance of getting COVID. This may also reflect the translation of “may” to “can possibly” in the Norwegian summary.

Based on the results of the U.S. trial, using plain language increased the proportion of participants who agreed or strongly agreed that not enough is known to be sure about the effects of wearing glasses to reduce the chance of getting COVID from 67.4% to 88.9% (risk difference 21.4%, 95% CI 9.8 to 33.1) (Table 3b). The effect of using GRADE language was less certain (Table 3b).

Table 3. Risk differences for sufficiency of the evidence

3a. Norwegian Trial

Not enough is known to be sure about the effects of wearing glasses to reduce the chance of getting COVID

Strongly agree (correct)
Agree (correct)
Disagree
Strongly disagree

Overall uncertainty	Margin of Error	n/N (%)[a]	Odds ratio [95% CI][b]	Risk difference [95% CI][c]	*P*-value
No explicit language	Not shown	65/80 (81.2%)	1	0	NA
No explicit language	Shown	59/71 (83.1%)	1.13 [0.49 to 2.62]	1.85 [-10.36 to 14.06]	.767
Plain language	Not shown	62/76 (81.6%)	1.02 [0.46 to 2.29]	0.33 [-11.88 to 12.54]	.958
Plain language	Shown	71/79 (89.9%)	2.05 [0.81 to 5.15]	8.62 [-2.21 to 19.46]	.127
GRADE language	Not shown	56/71 (78.9%)	0.86 [0.39 to 1.92]	-2.38 [-15.16 to 10.40]	.715
GRADE language	Shown	66/75 (88.0%)	1.69 [0.69 to 4.14]	6.75 [-4.53 to 18.03]	.249

[a]Number (n) of participants randomized to the intervention (N) answering correctly or as anticipated. [b]Odds ratios include the main and interaction effects. [c]Risk differences account for uncertainty on the baseline odds..

3b. U.S. trial

Not enough is known to be sure about the effects of wearing glasses to reduce the chance of getting COVID

Strongly agree (correct)
Agree (correct)
Disagree
Strongly disagree

Overall uncertainty	Margin of Error	n/N (%)[a]	Odds ratio [95% CI][b]	Risk difference [95% CI][c]	*P*-value
No explicit language	Not shown	58/86 (67.4%)	1	0	NA
No explicit language	Shown	59/88 (67.0%)	0.98 [0.52 to 1.85]	-0.40 [-14.34 to 13.55]	.956
Plain language	Not shown	88/99 (88.9%)	3.86 [1.78 to 8.36]	21.45 [9.77 to 33.13]	.001
Plain language	Shown	81/90 (90.0%)	4.34 [1.91 to 9.90]	22.56 [10.87 to 34.24]	.001
GRADE language	Not shown	63/79 (79.7%)	1.90 [0.93 to 3.87]	12.30 [-0.98 to 25.59]	.076
GRADE language	Shown	73/101 (72.3%)	1.26 [0.67 to 2.36]	4.84 [-8.37 to 18.04]	.472

[a]Number (n) of participants randomized to the intervention (N) answering correctly or as anticipated. [b]Odds ratios include the main and interaction effects. [c]Risk differences account for uncertainty on the baseline odds..

Understanding of the certainty of important harms

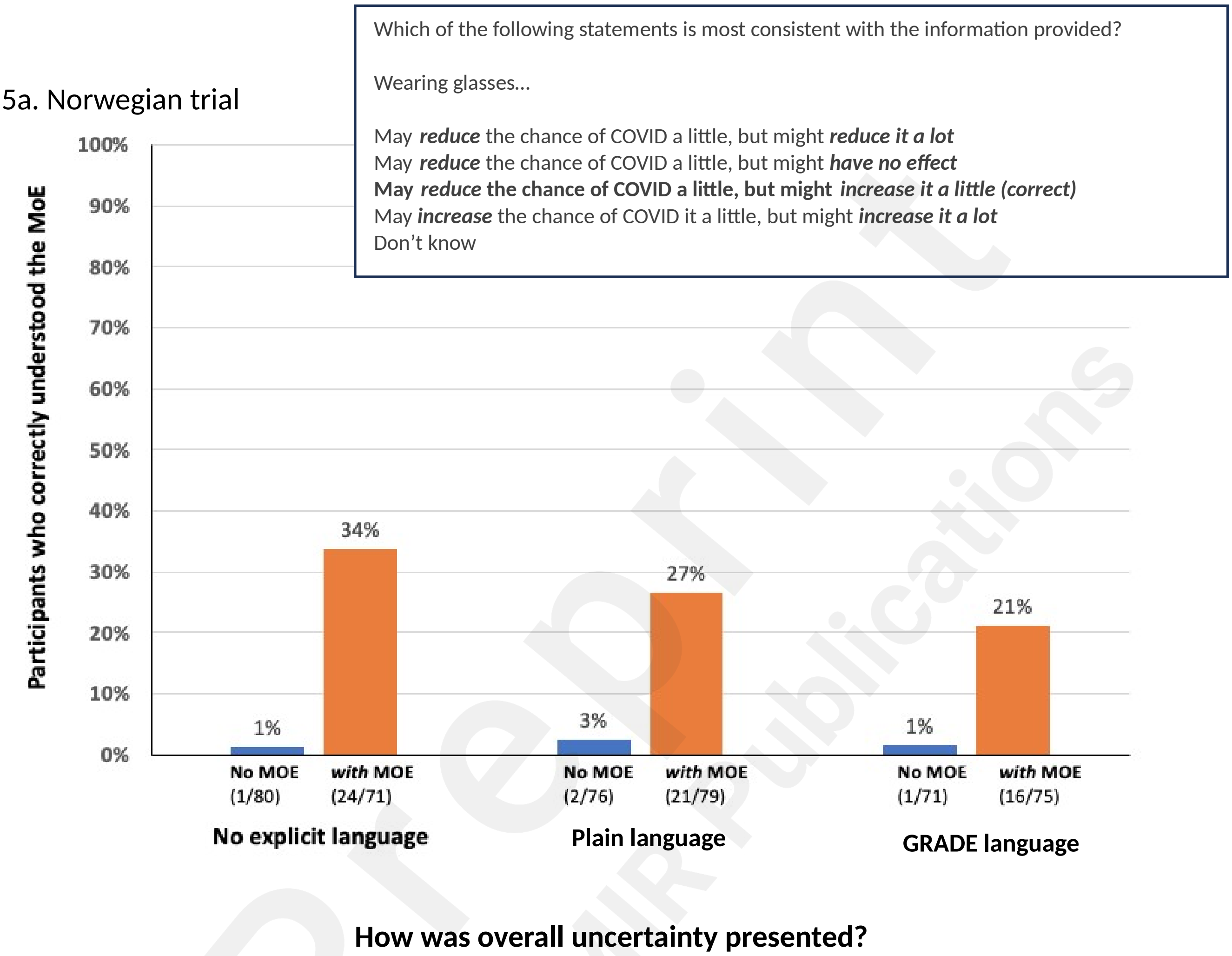
Individual trials

Understanding of the certainty of important harms varied across the trials. About 14% of participants in the Norwegian trial and from 28.7% to 35.6% in the U.S. trial correctly understood what we aimed to communicate about the certainty of the evidence for serious harms (Appendix 5, Tables 5a and 5b). The difference between the Norwegian and U.S. trials might, again, have been due to translation. The English version referred to "your chance of any major harms, such as a serious fall due to reduced vision." Whereas the Norwegian version, if back translated, referred to "the risk of serious injury, e.g. after falling" and did not include "due to reduced vision." Nonetheless, in both trials, there were only small differences between the comparison groups that could have occurred by chance alone.

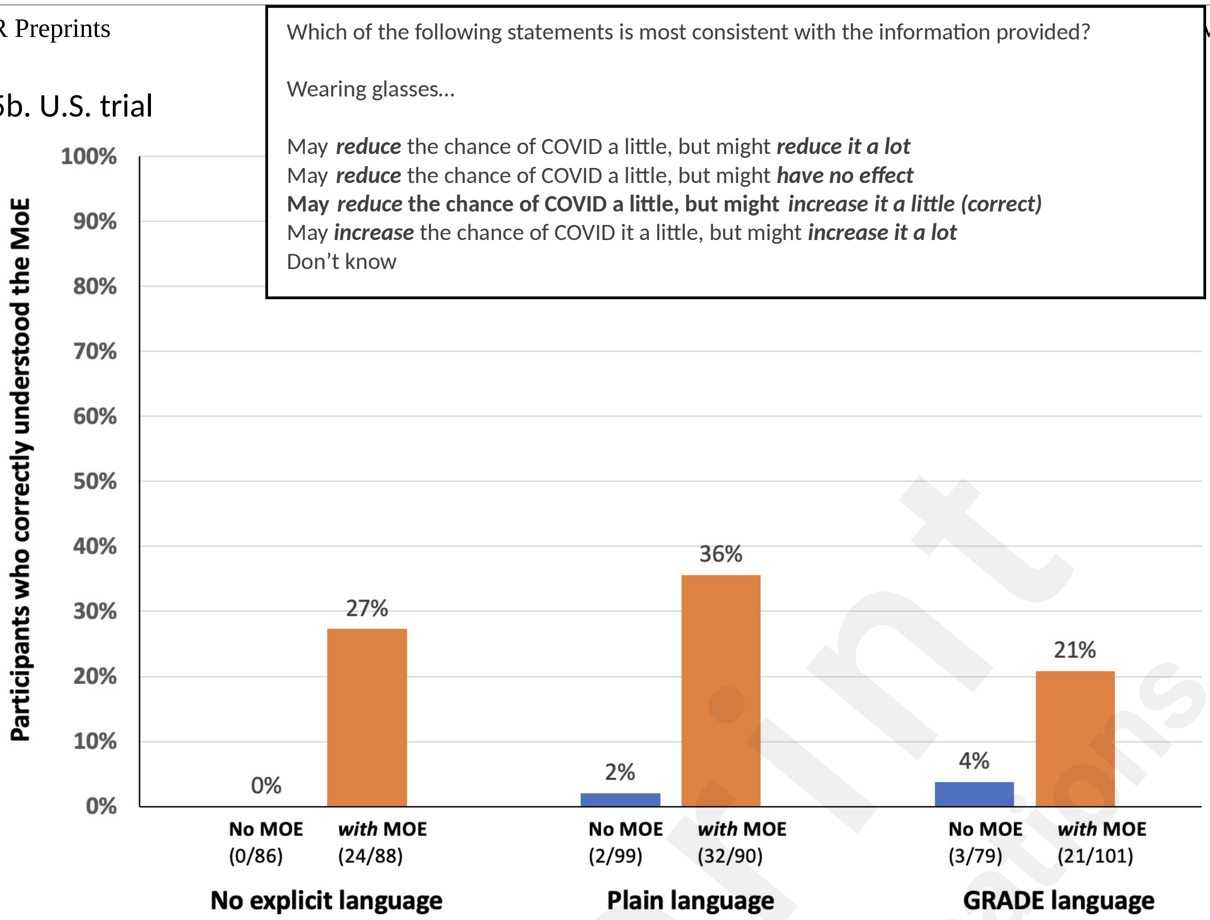
Reporting or not reporting the margin of error

Showing compared to not showing the margin of error increased the proportion of people who (correctly) chose the correct answer most consistent with the information provided: Wearing glasses may reduce the chance of COVID a little but might increase it a little (e.g., 21% to 34% vs. 1% to 3%; and 21% to 36% vs. 0% to 4%, in the Norwegian and U.S. trials, respectively) (Figures 5a and 5b; also Appendix 6, Tables 6a and 6b).

Figure 5. Understanding the Margin of Error. Effect of "margin of error" with different ways of presenting *overall uncertainty* in the Norwegian (A) and U.S. (B) trials. The figure shows the percentage of participants *correctly* answering the question in the box (correct response shown in bold).



5b. U.S. trial


Which of the following statements is most consistent with the information provided?
Wearing glasses…
May reduce the chance of COVID a little, but might reduce it a lot
May reduce the chance of COVID a little, but might have no effect
May reduce the chance of COVID a little, but might increase it a little (correct)
May increase the chance of COVID it a little, but might increase it a lot
Don't know
Participants who correctly understood the MoE
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
27%
2%
36%
4%
21%
No MOE (0/86)
with MOE (24/88)
No MOE (2/99)
with MOE (32/90)
No MOE (3/79)
with MOE (21/101)
No explicit language
Plain language
GRADE language

Nevertheless, this means most participants 64% to 79%) *failed* to correctly understand the margin of error when it was shown. Unsurprisingly, very few people correctly guessed the margin of error when it was not shown. A correct answer when not shown the margin of error would have been "Don't know." The proportion of participants who responded "don't know" when the margin of error was not shown was 2.3% compared to 2.4% when it was shown.

Secondary outcomes

Interest in wearing glasses to reduce Covid risk during a surge in cases

Interest in wearing glasses to reduce risk during a surge was consistently lower in the Norwegian vs. U.S. trials (Appendix 7).

In the Norwegian trial, plain language with or without the margin of error and GRADE language without the margin of error reduced the proportion of participants who responded that they would be likely or very likely to wear or recommend wearing glasses to reduce the chance of getting COVID if there were a surge of COVID cases (risk difference -13.6%, -24.7 to -2.5; -18.6%, 95% CI -28.7 to -8.4; -15.5%, -26.4 to -4.5 respectively) (Appendix 8, Table 8a).

In the U.S. trial, the combination of using either plain or GRADE language to communicate the overall uncertainty of the benefit of wearing glasses and showing the margin of error reduced the proportion of participants who responded that they would be likely or very likely to wear or recommend wearing glasses to reduce the chance of getting COVID if there were a surge of COVID cases (risk difference -20.5%, 95% CI -33.1 to -7.9 and 21.2%, 95% CI -33.5 to -8.9 respectively) (Appendix 8, Table 8b).

The difference between the two trials is due in part to the lower proportion of participants in the Norwegian reference group compared to the U.S. reference group that responded that they would be likely or very likely to wear or recommend wearing glasses to reduce the chance of getting COVID if there were a surge of COVID cases (22.5% and 36.0% respectively). This, again, might be due to the translation from English to Norwegian.

Interest in wearing glasses to reduce Covid risk if no surge in cases

We did not find an effect of plain or Grade language compared to no explicit language on the likelihood of wearing glasses or recommending wearing glasses to reduce the chance of getting COVID if there were very few COVID cases, the perceived benefit of wearing glasses, or the perceived chance of important harms (Appendix 9).

Perceptions of information provided

In both trials, perceptions of information provided differed little across plain, GRADE compared to no language, with few exceptions:

In both trials (US more so than Norway), perceptions of *helpfulness* were somewhat lower with plain language, with or without margin of error (Appendix 10, Table 10a and 10b).

In the Norwegian trial, with plain or GRADE language, reporting margin of error reduced perceptions of the *trustworthiness* of the information. (Appendix 11, Table 11a).

In the Norwegian trial, GRADE language without the margin of error reduced the perception that information about whether wearing glasses affects the chance of getting COVID was *sufficient* (Appendix 12, table 12a).

In the Norwegian trial, plain language without the margin of error and GRADE language with it reduced perceptions that the information about the benefit of glasses was *clear.* A similar effect was seen in the US trial for plain language with the margin of error (Appendix 13, Table 13a and 13b).

In the Norwegian trial, plain language with the margin of error reduced perceptions that information about whether glasses had important harms was *clear*. This effect was seen in the US trial showing the margin of error with no explicit or plain language. (Appendix 14, table 14a and 14b).

In the Norwegian trial, plain language with the margin of error reduced the proportion of participants who responded that they definitely or probably would *share the information* with someone who heard that wearing glasses might affect your chance of getting COVID (Appendix 15, Table 15a).

Decisional conflict

Plain language without the margin of error increased the feeling that the decision about wearing glasses if there were a surge of COVID cases was hard to make (Appendix 16, Table 16a). Most participants felt that they made an informed decision about wearing glasses if there were a surge of COVID (52.6 to 67.6% in the Norwegian trial and 70.9 to 83.0% in the U.S trial) (Appendix 17, Tables 17a and 17b).

Potential modifying factors

We did not find credible evidence of effect modification for numeracy, education, salience, or the time taken to complete the questionnaire.

Discussion

The U.S. and Norwegian trials comparing different ways of communicating the overall and statistical uncertainty of research results to the public generated mixed results. Plain language improved readers' understanding of overall uncertainty of the benefit, but only to a modest extent. The effect of GRADE language was uncertain, but at best had a modest effect. Also reporting the margin of error reduced understanding, by making the evidence seem more uncertain than it actually was. Reporting the margin of error did improve understanding of statistical uncertainty around the effect of glasses, but only for a minority of people.

Communicating overall uncertainty of the benefit

Plain language is probably more accessible than GRADE language (e.g., the phrase "not very confident" seems easier to understand than "moderate uncertainty") for communicating the overall uncertainty of benefit. Given the modest findings, more work is needed on the language, other ways of communicating uncertainty, such as visualizations, and easily accessible explanations.

It should be noted that we displayed the GRADE symbol together with GRADE language. The GRADE symbols for uncertainty are similar to widely used symbols used for ranking the quality of, for example, hotels, restaurants, and consumer products. [30] Nonetheless, they may be unfamiliar in this context and not easily understood without further explanation.

In another small trial [31] comparing GRADE symbols to letters to convey the quality of evidence, both letters and symbols were well understood.

Both the plain language and GRADE language summaries included text explaining why we were “not very confident” (it was “uncertain” that wearing glasses may slightly reduce your chance of getting Covid). These explanations were available under a tab labelled “Keep in mind…”. Neither the number of participants who read this text, nor the effect of the text on their understanding of the certainty of the evidence are known. Although we user-tested the summaries prior to the trials, further exploration of the extent to which people find this text helpful and how it impacts their understanding of the uncertainty of the evidence is warranted.

Communicating overall uncertainty of serious harm

Neither plain language nor GRADE language improved the understanding of overall uncertainty of serious harm. This could be because we did not explain why we said there was moderate certainty evidence, and prima facie, it seemed implausible to participants that wearing glasses could cause serious harm. In fact, about one third of the Norwegian participants were very sure, and this might be because they considered it implausible that wearing glasses could cause serious harms. On the other hand, other participants might have been somewhat or very unsure because there was so little information (about half).

Roughly twice as many participants in the U.S. trial compared to participants in the Norwegian trial correctly understood the overall uncertainty of the evidence for serious harm. The most likely reason for this, again, is a difference between the English summaries and the Norwegian translations. In this case, the explanation for why serious harm (serious injury after falling is plausible (because of reduced vision) was left out of the Norwegian translation. This again highlights the need for more extensive user testing and the use of back translation or other means of ensuring that translations are correctly understood.

Effect of margin of error on overall uncertainty

Showing the MOE did help people understand statistical uncertainty (i.e., glasses may reduce the chance of covid a little *but might* increase it a little). While this is encouraging, the modest effect means more work is needed to ensure that people understand what MOE is getting at.

In the group of participants shown plain language, showing the margin of error (for benefit) decreased the proportion of people who answered this question correctly. This probably happened because the effect of glasses was small, and MOE really highlighted this: adding uncertainty to uncertainty in context of small effect to begin with made people feel the effect was less certain than it actually was. This finding underscores the need for more work to help people calibrate their sense of uncertainty. It would be interesting to see what would happen in another example where the intervention effect was bigger than the effect of glasses. Another reason for this may be that participants who were shown margin of error were also shown two reasons for rating the evidence as low certainty (wide margin of error *and* important study limitations) whereas those who were not shown margin of error were only presented with one source of uncertainty (important study limitations). Findings from a 2020 study looking at the effect of communicating uncertainty found that participants who were shown three sources of uncertainty were more likely to report a weaker perception of the effectiveness of the intervention (drug) than those who were presented with only two sources of uncertainty. [32] More research needs to be conducted to explore this hypothesis.

Effect margin of error on intended behavior

Reporting overall uncertainty using plain language or GRADE language or reporting the margin of error decreased the likelihood that participants would wear glasses to protect against COVID if there was a surge in cases. This finding is consistent with other research finding that reporting uncertainty decreased the likelihood that people would use eye protection to reduce their chance of getting COVID. [17] It is also logical that the less sure you are of the benefits of doing something, the less likely you are to do it. These results highlight the importance of effectively communicating uncertainty if the intention is to inform people rather than to persuade them. [33]

Effect on margin of error on perceptions of information

The Norwegian trial, showing the margin of error in combination with plain language reduced the perceived trustworthiness of the information (the results were similar though smaller, and not statistically significant for the other versions). This is consistent with the findings of Schneider et al., [17] which found that including a clue that evidence of the effect of eye protection was low quality/certainty reduced perceived trustworthiness of an infographic. In contrast, in our U.S. trial, the reduction in perceived trustworthiness when showing the margin of error was substantially smaller and not statistically significant in any version. This suggests we cannot conclude that communicating uncertainty necessarily reduces the perceived trustworthiness of information about the effects of interventions. Indeed, more candid communication over time might make changes in recommendations seem less arbitrary and help preserve people’s trust in health authorities. [33]

There is some evidence that plain language with or without the MoE and GRADE language with the MoE reduced the perception that the summary was sufficient. This may be due to participants confusing the sufficiency of the evidence (i.e., a small, uncertain effect) with sufficiency of the summary which was meant to communicate the effect. While the question we asked specifically aimed at the latter (i.e., whether the information was a sufficient summary of what is known about the effects of wearing glasses to reduce the chance of getting COVID), future qualitative work should be done to be help distinguish these two kinds of sufficiency. Similarly, showing MOE also reduced the perception that the summaries were clear, perhaps for the same reasons.

Nevertheless, most participants found the decision about wearing glasses to reduce the chance of getting COVID hard to make, with or without the MoE and regardless of which summary they were shown.

Our study has several strengths including the randomized factorial design, which let U.S. explore the interaction between uncertainty language (GRADE, plain or none) and the margin of error on understanding of understanding of benefits, harms and the corresponding uncertainties, and replication in 2 distinct populations. Study limitations include the weak, uncertain effect that we were trying to summarize (from the Glasses trial) which magnified the communication challenges, and several language translation issues which may limit the across country comparisons. Furthermore, our results may have been influenced by lack of saliency – i.e., that Covid infection rates were relatively low when we conducted the study and participants may respond differently in a more realistic/pressing scenario. We also need to explore the effect of paying participants in these online trials on the quality of the responses. [23]

Our study shows that explicitly reporting uncertainty affects peoples' understanding, perceptions and (intended) actions. We found that plain language was better than no explicit language in helping people understand overall certainty of evidence, though most participants still did not correctly understand how sure they could be. Reporting margin of error reduced understanding of overall uncertainty by making people feel the evidence was even less certain. Reporting margin of error improved interpretation of statistical uncertainty around the effect of glasses, but only for a minority of participants.

This study underscores how much more work needs to be done to develop effective ways to communicate overall uncertainty (*how confident can I be about this claim*?) and just how fuzzy the numbers are (statistical uncertainty). If this communication is done poorly, it may simply add to confusion and lead to poor decisions. Done well, effective communication around uncertainty can help people to make the best decisions they can, given the evidence that is known.

Acknowledgments

We thank Ingeborg Hess Elgersma and the CEIR team for their valuable feedback on the translations and questionnaire development.

Conflicts of Interest

None

References

1. Fischhoff B. The COVID Communication Breakdown: How to Fix Public Health Messaging. Foreign Affairs. October 2021. https://www.foreignaffairs.com/articles/united-states/2021-10-04/covid-communication-breakdown.
2. Fischhoff B. The sciences of science communication. *Proc Natl Acad Sci* 2013;110 (suppl 3):14033-14039.
3. Fischhoff B, Krishnamurti T. Risk perception and communication. In: *Oxford Textbook of Global Public Health. 7th ed.* Oxford: Oxford University Press; 2022.
4. Oxman AD GC, Flottorp S, Lewin S, Rosenbaum S, Fretheim A. Development of a checklist for people communicating evidence-based information about the effects of healthcare interventions: a mixed methods study.*BMJ Open.* 2020;10:e036348. https://doi.org/036310.031136/bmjopen-032019-036348.
5. Woloshin S, Yang Y, Fischhoff B. Communicating health information with visual displays. Nat Med 29, 1085–1091 (2023). https://doi.org/10.1038/s41591-023-02328-1.
6. Woloshin S, Dewitt B, Krishnamurti T, Fischhoff B. Assessing How Consumers Interpret and Act on Results From At-Home COVID-19 Self-test Kits: A Randomized Clinical Trial. *JAMA Intern Med.* 2022;182:332–341.
7. Woloshin S, Schwartz L. Communicating data about the benefits and harms of treatment: A randomized trial. *Ann Intern Med* 2011 155:87-96.
8. Akl E, Oxman A, Herrin J, et al. Using alternative statistical formats for presenting risks and risk reductions.Cochrane Database of Systematic Reviews 2011, Issue 3. Art. No.: CD006776.DOI: 10.1002/14651858.CD006776.pub2.
9. Carling CL, Kristoffersen DT, Montori VM, Herrin J, Schünemann HJ, Treweek S, Akl EA, Oxman AD. The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomized trial. PLoS Med. 2009 Aug;6(8):e1000134. doi: 10.1371/journal.pmed.1000134.
10. Schwartz LM,Woloshin S. The Drug Facts Box: Improving the communication of prescription drug information. Proc Natl Acad Sci U S A. 2013;110(suppl 3):14069-14074. doi:10.1073/pnas.1214646110.
11. Schwartz LM, Woloshin S. Communicating uncertainties about prescription drugs to the public: a national randomized trial. Arch Intern Med. 2011 Sep 12;171(16):1463-8. doi: 10.1001/archinternmed.2011.396. PMID: 21911629.
12. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401-6. https://doi.org/10.1016/j.jclinepi.2010.07.015.
13. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011;64(12):1283-93. https://doi.org/10.1016/j.jclinepi.2011.01.012.
14. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ, et al. What is "quality of evidence" and why is it important to clinicians? BMJ. 2008;336(7651):995-8.

https://doi.org/10.1136/bmj.39490.551019.be.

15. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4.mRating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407- 15. https://doi.org/10.1016/j.jclinepi.2010.07.017.
16. Santesso N, Rader T, Nilsen ES, Glenton C, Rosenbaum S, Ciapponi A, et al. (2015) A summary to communicate evidence from systematic reviews to the public improved understanding and accessibility of information: a randomized controlled trial. J Clin Epidemiol 68:182–90. https://doi.org/10.1016/j. jclinepi.2014.04.009 PMID: 25034199.
17. Schneider CR, Freeman ALJ, Spiegelhalter D, van der Linden S (2021) The effects of quality of evidence communication on perception of public health information about COVID-19: Two randomised controlled trials. PLoS ONE 16(11): e0259048. https://doi.org/10.1371/journal.pone.0259048.
18. Fretheim A, Elgersma IH, Helleve A, Elstrøm P, Kacelnik O, Hemkens LG. Glasses Against transmission of SARS-CoV-2 in the community (GLASSY): a pragmatic randomized trial.medRxiv. 2022:2022.07.31.22278223. https://doi.org/10.1101/2022.07.31.22278223.
19. Oxman AD, Fretheim A, Lewin S, Flottorp S, Glenton C, Helleve A, et al. Health communicationin and out of public health emergencies: to persuade or to inform? Health Res Policy Syst. 2022;20:28. https://doi.org/10.1186/s12961-022-00828-z.
20. Prolific. Easily find vetted research participants and AI taskers at scale. https://www.prolific.com/ (accessed 15 March 2024).
21. Opinion. Vi måler nordmenns meninger, og leverer innsikt som bringer deg videre. https://www.opinion.no/ (accessed 15 March 2024).
22. Bradley P. Bots and data quality on crowdsourcing platforms (2018). https://www.prolific.com/resources/bots-and-data-quality-on-crowdsourcing-platforms (accessed 20 February 2024).
23. Douglas BD, Ewell PJ, Brauer M. Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. PLoS ONE 18(3): e0279720. https://doi.org/10.1371/journal.pone.0279720.
24. ISO. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. Geneva; 2019.
25. Loop11.Empower Your Team: Data-driven Design Decisions Made Easy. https://www.loop11.com/ (accessed 15 March 2024).
26. O'Connor AM. Validation of a decisional conflict scale. Med Decis Making. 1995;15(1):25-30.
27. Woloshin, S., Oxman, A. D., Rose, C. J., Rosenbaum, S. E., Holst, C., Diaz-Hemard, L., Rada, G., Vergara, C., & Munthe-Kaas, H. (2022). Communicating uncertainty about the effects of wearing glasses to reduce the chance of getting COVID: protocol for a randomized trial (Version 8). Zenodo. https://doi.org/10.5281/zenodo.7428981.
28. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997 Dec 1;127(11):966-72. doi: 10.7326/0003-4819-127-11-199712010-00003. PMID: 9412301.
29. Formuleringer i GRADE for å uttrykke resultater fra kunnskapsoppsummeringer om effekt" - "LAV tillit til kunnskapsgrunnlaget. https://www.fhi.no/ku/oppsummert-forskning-for-helsetjenesten/metodeboka/framgangsmate/hente-ut-data-sammenfatte-og-gradere/?term=

(accessed 15 March 2024).

30. Schunemann HJ, Best D, Vist G, Oxman AD, Group GW. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. CMAJ. 2003;169(7):677-80. http://www.ncbi.nlm.nih.gov/pmc/articles/pmc202287/

31. Akl EA, Maroun N, Guyatt G, Oxman AD, Alonso-Coello P, Vist GE, et al. Symbols were superior to numbers for presenting strength of recommendations to health care consumers: a randomized trial. J Clin Epidemiol. 2007;60(12):1298-305. https://doi.org/10.1016/j.jclinepi.2007.03.011

32. Büchter, R. B., Betsch, C., Ehrlich, M., Fechtelpeter, D., Grouven, U., Keller, S, Waltering, A. Communicating uncertainty in written consumer health information to the public: parallel-group, web-based randomized controlled trial. J Med Internet Res. 2020;22(8):e15899.

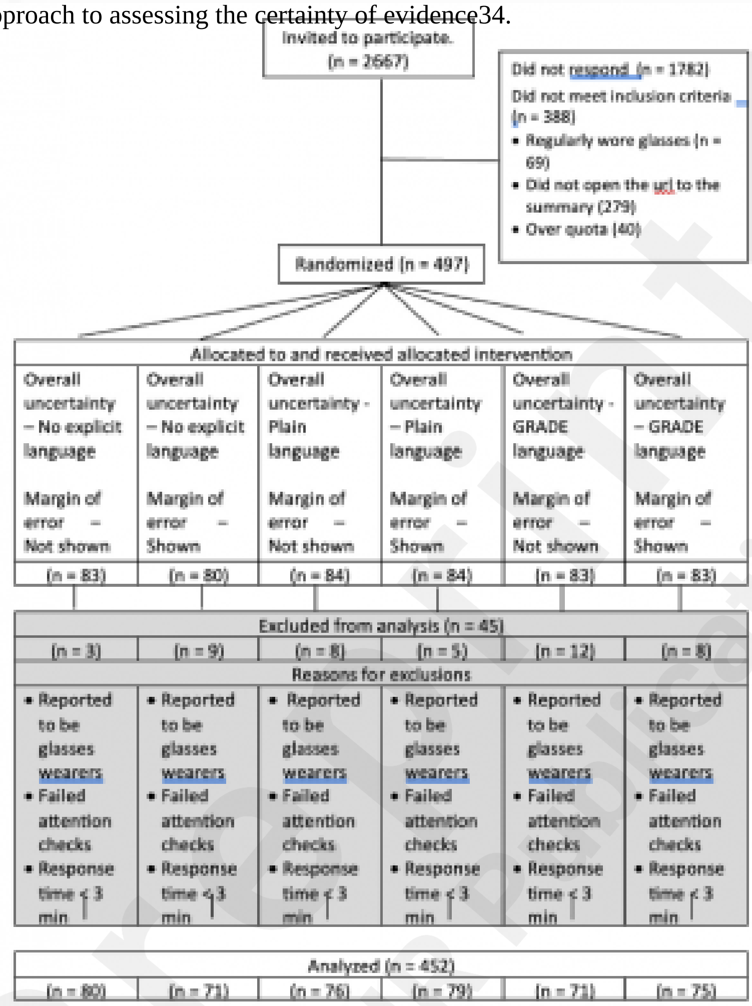
33. Oxman AD, Fretheim A, Lewin S, Flottorp S, Glenton C, Helleve A, et al. Health communication in and out of public health emergencies: to persuade or to inform? Health Res Policy Syst. 2022;20:28. https://doi.org/10.1186/s12961-022-00828-z.

34. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011;64(12):1283-93. https://doi.org/10.1016/j.jclinepi.2011.01.012.

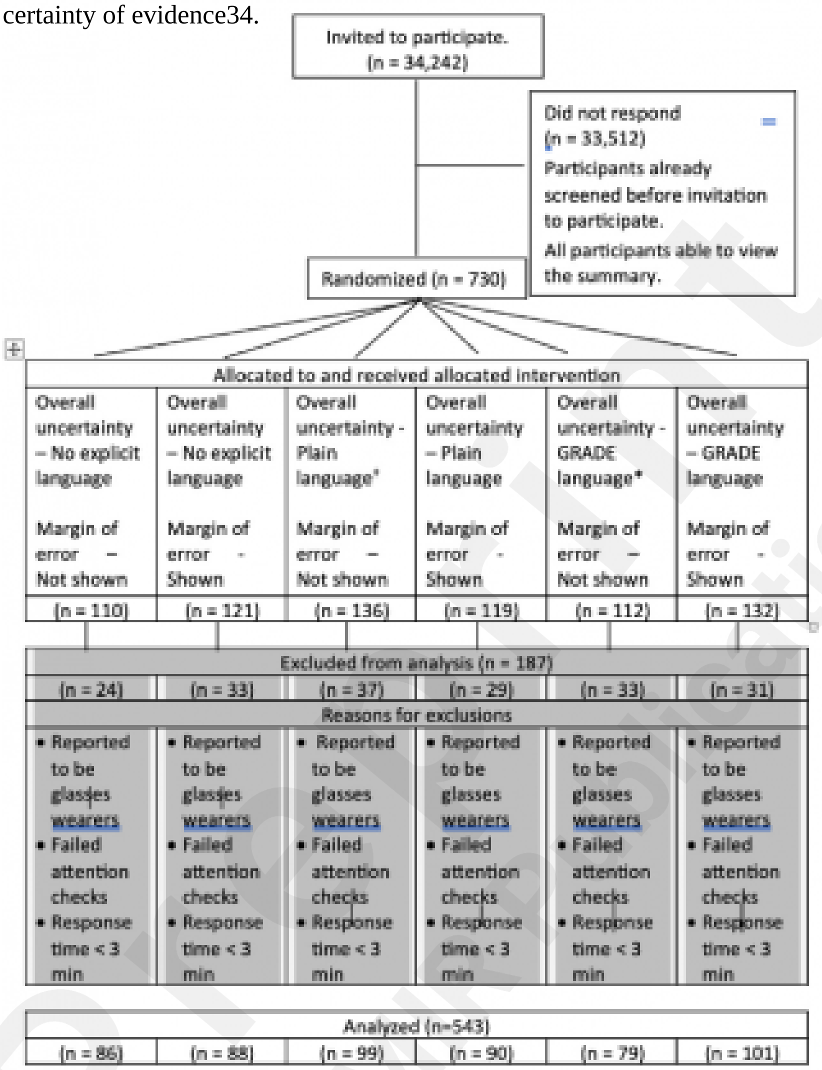
Supplementary Files

Figures

Flow diagram of study participants in the Norwegian trial **there are table footnotes too. Here they are: aA less formal expression of the overall uncertainty language used in ordinary or familiar conversation, corresponding to the same GRADE assessment of the certainty of the evidence. bBased on the Cochrane Effective Practice and Organisation of Care Group's guidance for communicating the certainty of evidence based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to assessing the certainty of evidence34.



Flow diagram of study participants in the U.S. trial figure has footnotes: aA less formal expression of the overall uncertainty language used in ordinary or familiar conversation, corresponding to the same GRADE assessment of the certainty of the evidence. bBased on the Cochrane Effective Practice and Organisation of Care Group's guidance for communicating the certainty of evidence based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to assessing the certainty of evidence34.



A. Overall uncertainty language about the possible benefit of glasses to reduce Covid used in each intervention group in the U.S. trial (the Norwegian version is in Appendix). B. Margin of error language was presented in the same way in the 3 versions reporting it.

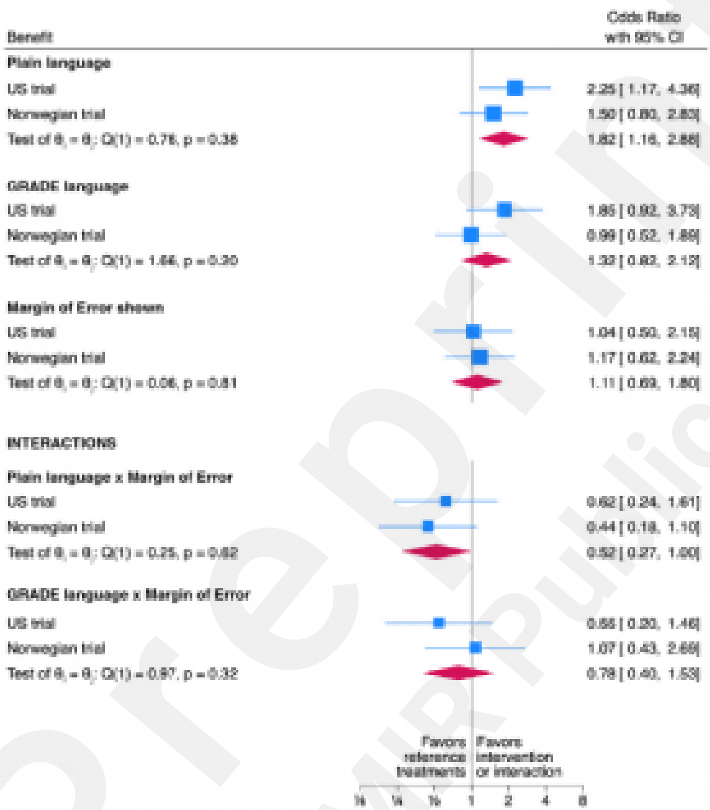
(a) Overall uncertainty

	What are the effects of wearing glasses to reduce the chance of getting COVID?
GRADE language*	What is the benefit? Wearing glasses may slightly reduce your chance of getting COVID (⊕⊕◯◯ low certainty evidence). What are the downsides? Wearing glasses probably does not increase your chance of any major harms, such as a serious fall due to reduced vision (⊕⊕⊕◯ moderate certainty evidence). But some people are irritated by foggy glasses, and some feel silly wearing them.
Plain language	What is the benefit? Wearing glasses may reduce your chance of getting COVID a little – but we are not very confident about this. What are the downsides? We are somewhat confident that wearing glasses does not cause important harms such as a serious fall due to reduced vision. But some people are irritated by foggy glasses, and some feel silly wearing them.
No explicit language	What is the benefit? Wearing glasses may reduce your chance of getting COVID a little. What are the downsides? Wearing glasses probably does not cause important harms such as a serious fall due to reduced vision. But some people are irritated by foggy glasses, and some feel silly wearing them.

(b) *Margin of error*

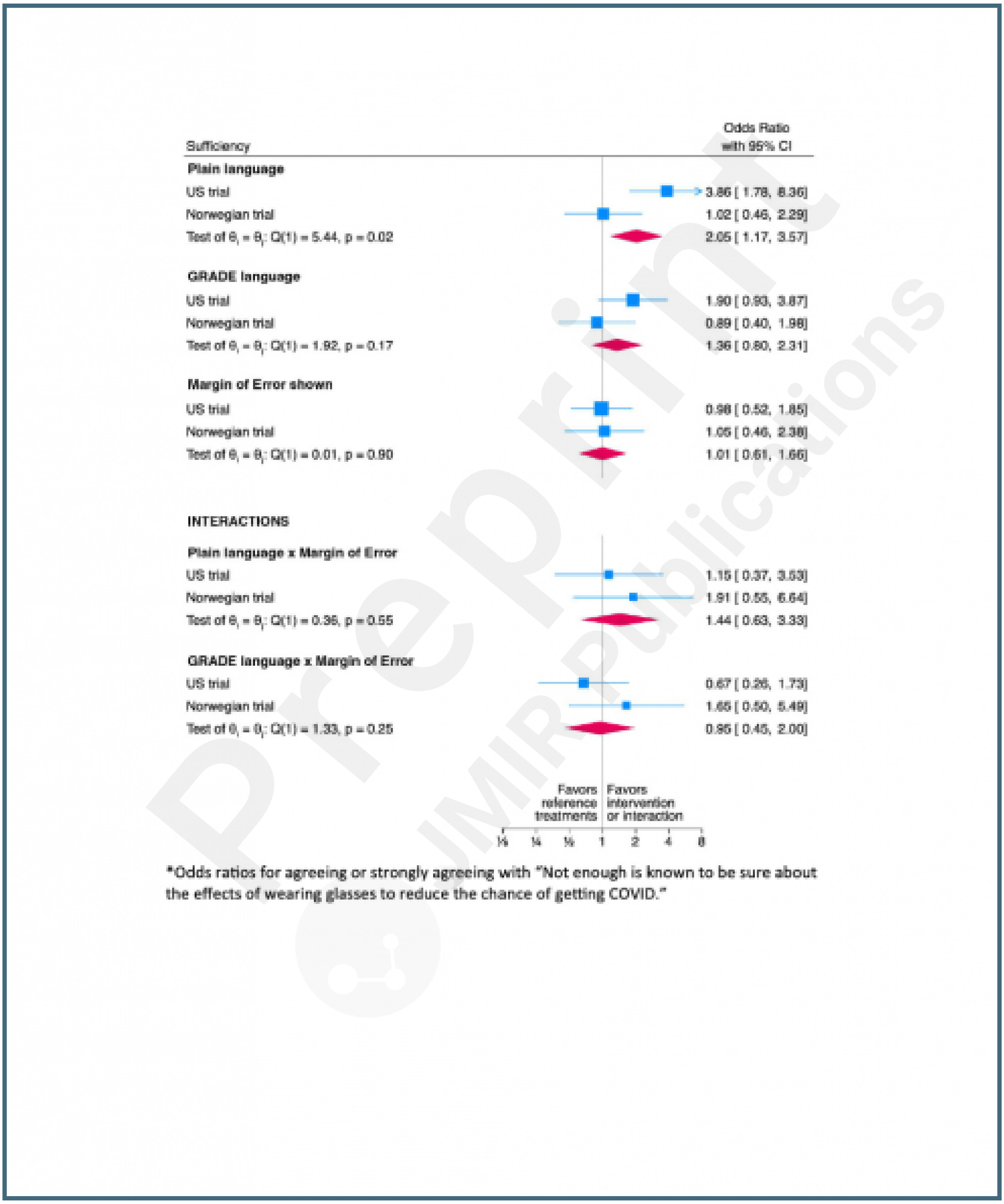
	Here's what happened:
Reported	Over the next 2 weeks, slightly fewer people told to wear glasses tested positive for COVID: 9.6% compared to 11.5% of those told not to wear glasses. That's a difference of 1.9%. Accounting for the play of chance (i.e., the margin of error), glasses might *reduce* the risk of COVID by as much as 3.9% but might *increase* it by as much as 0.1%.
Not reported	Over the next 2 weeks, slightly fewer people told to wear glasses tested positive for COVID: 9.6% compared to 11.5% of those told not to wear glasses. That's a difference of 1.9%.

Understanding of the certainty of the benefit of wearing glasses* There is a footnote: *NOTE: Odds ratios for answering “Mixed but more unsure than sure” to the question “How sure are you about the effect of wearing glasses on your chance of getting COVID?”.

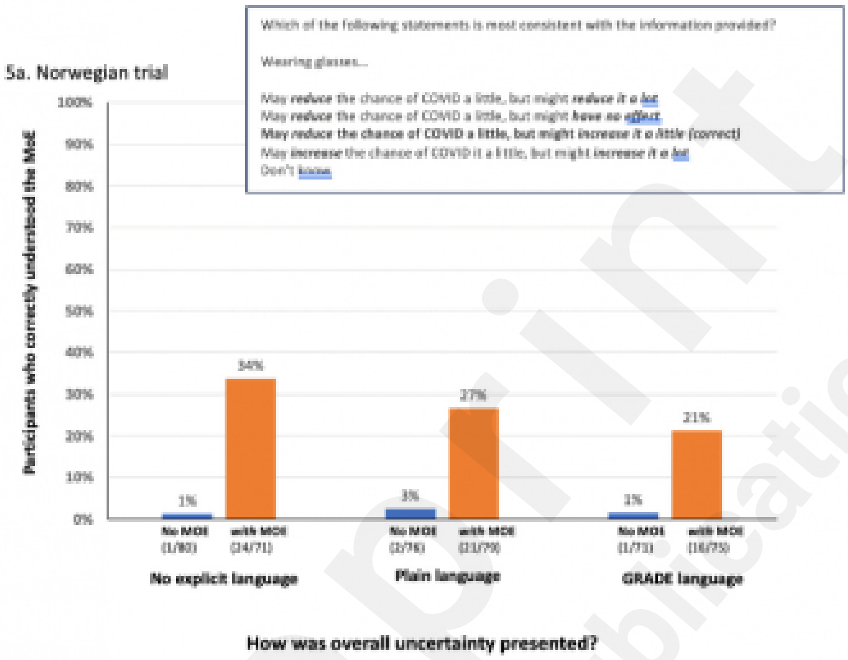


*Odds ratios for answering “Mixed but more unsure than sure” to the question “How sure are you about the effect of wearing glasses on your chance of getting COVID?”.

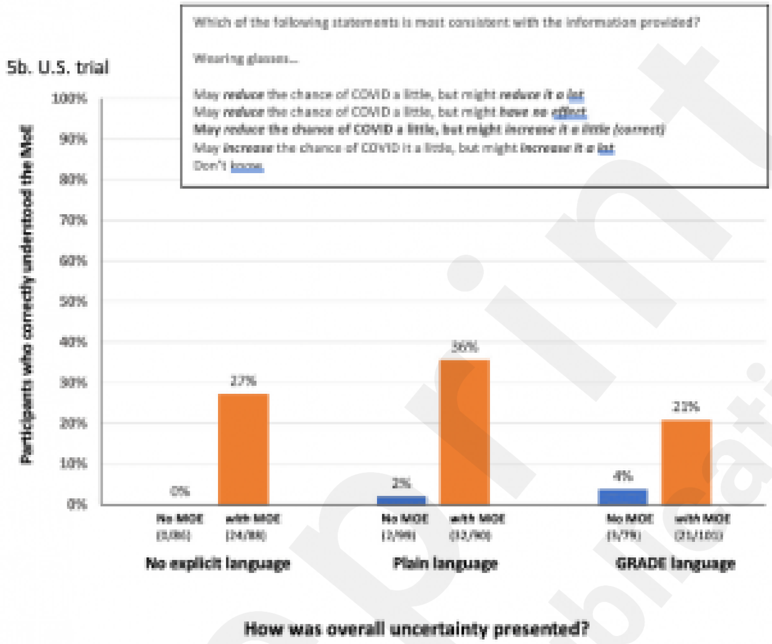
Understanding sufficiency of the evidence* There is a footnote: NOTE: Odds ratios for agreeing or strongly agreeing with “Not enough is known to be sure about the effects of wearing glasses to reduce the chance of getting COVID.”.



Understanding the Margin of Error. Effect of "margin of error" with different ways of presenting overall uncertainty in the Norwegian trial. The figure shows the percentage of participants correctly answering the question in the box (correct response shown in bold).



Understanding the Margin of Error. Effect of "margin of error" with different ways of presenting overall uncertainty in the US trial. The figure shows the percentage of participants correctly answering the question in the box (correct response shown in bold).



Multimedia Appendixes

Supplementary materials.
URL: http://asset.jmir.pub/assets/1c92d82a5ef01743bb29ba9fecec3ba1.docx