BMJ Open

How do the UK public interpret COVID-19 test results? Comparing the impact of official information about results and reliability used in the UK, USA and New Zealand: a randomised controlled trial

Gabriel Recchia ,[1] Claudia R Schneider ,[1,2] Alexandra LJ Freeman [1]

¹Winton Centre for Risk & Evidence Communication, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, UK
²Department of Psychology, University of Cambridge, Cambridge, UK

Correspondence to
Dr Gabriel Recchia;
glr29@cam.ac.uk

ABSTRACT

Objectives To assess the effects of different official information on public interpretation of a personal COVID-19 PCR test result.

Design A 5×2 factorial, randomised, between-subjects experiment, comparing four wordings of information about the test result and a control arm of no additional information; for both positive and negative test results.

Setting Online experiment using recruitment platform Respondi.

Participants UK participants (n=1744, after a pilot of n=1657) quota-sampled to be proportional to the UK national population on age and sex.

Interventions Participants were given a hypothetical COVID-19 PCR test result for 'John' who was presented as having a 50% chance of having COVID-19 based on symptoms alone. Participants were randomised to receive either a positive or negative result for 'John', then randomised again to receive either no more information, or text information on the interpretation of COVID-19 test results copied in September 2020 from the public websites of the UK's National Health Service, the USA's Centers for Disease Control, New Zealand's Ministry of Health or a modified version of the UK's wording. Information identifying the source of the wording was removed.

Main outcome measures Participants were asked 'What is your best guess as to the percent chance that John actually had COVID-19 at the time of his test, given his result?'; questions about their feelings of trustworthiness in the result, their perceptions of the quality of the underlying evidence and what action they felt 'John' should take in the light of his result.

Results Of those presented with a positive COVID-19 test result for 'John', the mean estimate of the probability that he had the virus was 73% (71.5%–74.5%); for those presented with a negative result, 38% (36.7%–40.0%). There was no main effect of information (wording) on these means. However, those participants given the official information from the UK website, which did not mention the possibility of false negatives or false positives, were more likely to give a categorical (100% or 0%) answer (UK: 68/343, 19.8% (15.9%–24.4%)); control group: 42/356,

Strengths and limitations of this study

► This study provides the first empirical evidence, to our knowledge, of how the public interpret the inherent uncertainty around COVID-19 test results, and information given alongside such results.

► The study was carried out on a large quota-sampled pool of UK participants, and tested wording taken from official websites in three countries about the interpretation of COVD-19 test results. However, the wording on these websites is likely different from the information provided to individuals immediately after their tests in the respective countries.

► A limitation of the study is that the prior probability of infection (based on symptoms/background prevalence) of the individual in our scenario was very high (50%), which would not commonly be the case. This was chosen as we felt it would not bias participants towards believing that the individual did or did not have COVID-19 before his test result was revealed, making the results easier to interpret.

► All behavioural responses from participants were purely hypothetical and framed as what the recipient of such a test result should do, so should be interpreted with caution.

11.8% (8.8%–15.6%)); the reverse was true for those viewing the New Zealand (NZ) wording, which highlighted the uncertainties most explicitly (20/345: 5.8% (3.7%–8.8%)). Aggregated across test result (positive/negative), there was a main effect of wording (p<0.001) on beliefs about how 'John' should behave, with those seeing the NZ wording marginally more likely to agree that 'John' should continue to self-isolate than those viewing the control or the UK wording. The proportion of participants who felt that a symptomatic individual who tests negative definitely should *not* self-isolate was highest among those viewing the UK wording (31/178, 17.4% (12.5%–23.7%)), and lowest among those viewing the NZ wording (6/159, 3.8% (1.6%–8.2%)). Although the NZ wording was rated harder to understand, participants reacted to the uncertainties

given in the text in the expected direction: there was a small main effect of wording on trust in the result (p=0.048), with people perceiving the test result as marginally less trustworthy after having read the NZ wording compared with the UK wording. Positive results were generally viewed as more trustworthy and as having higher quality of evidence than negative results (both p<0.001).

Conclusions The public's default assessment of the face value of both the positive and negative test results (control group) indicate an awareness that test results are not perfectly accurate. Compared with other messaging tested, participants shown the UK's 2020 wording about the interpretation of the test results appeared to interpret the results as more definitive than is warranted. Wording that acknowledges uncertainty can help people to have a more nuanced and realistic understanding of what a COVID-19 test result means, which supports decision making and behavioural response.

Preregistration and data repository Preregistration of pilot at osf.io/8n62f, preregistration of main experiment at osf.io/7rcj4, data and code available online (osf.io/pvhba).

INTRODUCTION

PCR tests for COVID-19 are widely used in many countries to help individuals take appropriate action such as self-quarantining and initiating contact-tracing. At a population level, they give policymakers vital information about the prevalence of the virus. However, the sensitivity and specificity of the tests (especially in real-world scenarios) are not 100%. Therefore, people taking action on the basis of test results need to be aware of the potential that for an individual, a single negative test may not be an 'all clear'. Similarly, it is important to be transparent that a positive test does not equate to a 100% chance that one carries the virus. Watson et al[1] created a tool to help with the interpretation of such tests. Using the approximate sensitivity and specificity values posited in Watson et al[1] on the basis of systematic reviews suggests that a positive test result with a prior suspicion of COVID-19 of 50% should be interpreted as around a 93% chance that the patient had COVID-19 at the time they were tested, and a negative test as around a 24% chance.[1] A recent assessment of the UK Office for National Statistics estimated a false positive rate of under 0.005%,[2] suggesting that even more confidence about the interpretation of a positive test result may be warranted.

However, the wording provided to support public interpretations of the test results varies from country to country. As of 2020, in the UK, the National Health Service (NHS) website on test interpretation expressed no uncertainty (e.g., 'A positive result means you had coronavirus when the test was done' and 'A negative result means the test did not find coronavirus').[3] On an equivalent US website from the Centers for Disease Control, the phrasing communicated slightly more uncertainty (e.g., 'If you test positive, know what protective steps to take to prevent others from getting sick' and 'If you test negative, you probably were not infected at the time your sample was collected. The test result only means that you did not have COVID-19 at the time of testing').[4,5] By contrast, in New Zealand (NZ), a much longer wording was provided by the Ministry of Health which included far more information about the uncertainties inherent in the test (e.g., 'A recent laboratory study found that different COVID-19 testing kits correctly detected COVID-19 in samples more than 95% (and frequently 100%) of the time. When tests were done on samples without the virus, the tests correctly gave a negative result 96% of the time. But it is important to remember that tests don't work as well in the real world.' and 'The viral test for COVID-19 is much better at correctly identifying people who don't have COVID-19 (this is known as a higher 'specificity'). We expect very few (if any) false positive test results (a false positive being a positive test result for someone who does not have the disease)'.)[6]

Concern has been expressed that information provided to patients without uncertainty may cause unwarranted confidence in the results, particularly negative results.[7] For example, if a negative test result is communicated as 'you do not have coronavirus', patients may take this statement at face value. However, if made aware that the results are sometimes wrong, they might become less certain about the meaning of the result. It is common to ask participants to provide subjective probabilities as a way of quantifying their uncertainty (both aleatory and epistemic).[8–11] It is important to note that such estimates should not be interpreted as calculations of real-world probabilities, or as readouts of probabilities assumed to exist 'in the head'; instead, they are likely to reflect an interplay of a range of psychological processes.[12–14] Nevertheless, there are good reasons to believe that such estimates do at least partially reflect participants' subjective beliefs[15] and do have predictive validity when used to assess likelihoods of future events.[9–11,16] In particular, the endpoints of the scale (100% and 0%) can be interpreted as extremely high certainty (that 'John' does or does not have the virus, respectively). The mid-point in English language surveys (50%) needs particular care in interpretation due to well-recognised lay usage of the term 'fifty-fifty' as a communication of 'I don't know',[17–19] but in any case, seems to communicate high levels of uncertainty. Our analyses therefore pay particular attention to these responses.

In addition to potentially affecting people's subjective beliefs about their COVID-19 status, decisions about whether and how to communicate uncertainty may also affect perceptions about the quality of evidence behind the information, and their level of trust in it.[20–27] For example, Johnson and Slovic[28] show that communication of uncertainty can have effects on perceived honesty and competence of the communicator. Broomell and Kane[29] report that perceptions of precision of the information, as a function of the communication of uncertainty, can influence people's perceived value of it. People's assessment of the information they are provided with in terms of credibility and confidence furthermore influences decision making and behaviour.[30,31] In the public health domain, Han[31] argues that 'people not only need but deserve information about uncertainty' (p15S), highlighting the importance of informed decision making

and the ethical principle of respect for patient autonomy. Investigating how the communication of uncertainty affects people's judgement of the information at hand and their behavioural intentions is thus critically important. In an experimental study in the context of COVID-19 public health communication, Wegwarth et al showed that participants preferred communication which explicitly stated uncertainty (in verbal and numeric format), both when asked how government and health experts should communicate, as well as when asked if the communication would motivate their compliance.[32] Finally, as alluded to above, there are ethical considerations to be weighed up around whether the goal should be to inform individuals so that they are well-positioned to take self-determined actions without coercion (see Blastland et al[33] for arguments in favour of this view), as opposed to persuading them to adopt a particular course of action. Note that more belief, trust and so on in a test result is not necessarily always good, even from the perspective of a public health official who wishes to persuade: if there is substantial uncertainty around a test's negative predictive value for individuals with symptoms, public health officials may well prefer symptomatic individuals with negative tests to exercise caution and self-isolate anyway.

This study set out to test the effects of different approaches to communicating COVID-19 PCR test results on how likely participants thought it was that the test recipient actually had COVID-19, on their perception of the quality of the evidence underlying the test result, the trustworthiness of the test result, and on participants'

views about whether positive/negative test result recipients should self-isolate. After a pilot study (see online supplemental information) we created an experimental version of the information ('experimental information' condition), based on the UK's NHS website wording but containing elements to make the uncertainties explicit. These elements were based loosely on the detailed uncertainty information provided on the New Zealand Ministry of Health website, but with less detail.

Preregistered hypotheses

We preregistered four hypotheses (osf.io/7rcj4) based on the findings of the pilot. Our specific research questions and corresponding hypotheses are summarised in table 1. In general, the pilot findings seemed to suggest a level of scepticism about positive test results after reading the UK or US information that was not present in other messaging conditions. We suspected this could be the case because the UK and US information did not explicitly note that positive test results were more reliable than negative test results, or indeed say much about the reliability of the tests at all (table 2), despite this being a hot topic in the news at the time. We reasoned that the absence of such information may have caused readers to be 'on their guard'.

METHODS

Participants aged 18+ were recruited for the main study online through the ISO-accredited company Respondi and directed

Table 1 Research questions and preregistered hypotheses

Research question	Pilot result	Formal hypothesis
Q1: What influence do message wording and test result have on how likely participants thought it was that the test recipient actually had COVID-19?	For participants in the 'positive' test result condition, estimates of the probability that 'John' actually had COVID-19 were lower for those shown the UK or US information than for those shown the New Zealand information or no information at all	H1: Interaction such that for participants in the 'positive' test result condition, **estimates of the probability that 'John' actually had COVID-19** will be lower for those shown the UK or US information than for those shown the New Zealand information, no information at all or experimental information based on the UK wording
Q2: What influence do message wording and test result have on participants' views of the quality of the evidence behind 'John's' test result?	Participants viewing no information perceived the quality of evidence behind the test to be higher if they had been shown positive (vs negative) test results; this difference was not found for participants viewing UK or US information	H2: Interaction such that participants viewing no information or experimental information who are shown *positive* test results will **perceive the quality of evidence** to be greater than individuals shown the same information and *negative* test results, whereas this difference will not be found for individuals shown US and UK information
Q3: What influence do message wording and test result have on participants' views of the trustworthiness of 'John's' test result?	Participants viewing no information and NZ information perceived the test result to be more trustworthy if they had been shown positive (vs negative) test results; this difference was not found for participants viewing UK information*	H3: Interaction such that participants viewing no information or experimental information who are shown *positive* test results will perceive the test result to be more **trustworthy** than individuals shown the same information and *negative* test results, whereas this difference will not be found for individuals shown US and UK information
Q4: What influence do message wording and test result have on participants' views about whether 'John' should self-isolate?	Pilot did not investigate participants' estimation of whether someone with a positive/negative result should self-isolate, but a hypothesis was generated on the basis of the results above	H4: Interaction such that participants viewing no information or experimental information who are shown *positive* test results will indicate greater **agreement with the statement that 'John' should self-isolate** than individuals shown the same information and *negative* test results, whereas this difference will not be found for individuals shown US and UK information

Dependent variables are in bold.

*The difference was found for US information, but with less confidence than for the New Zealand (NZ) and control conditions; due to the inconsistency with the other results, we assessed this as likely to be a false positive for the purposes of defining our formal hypotheses.

Table 2 Key characteristics of message wording from each country; full wording in online supplemental information

Short title	Original source*	Word count	Communicates possibility of false positive?	Communicates possibility of false negative?
US wording	US Centers for Disease Control and Prevention	225	No	Indirectly ('you probably were not infected at the time your sample was collected', with further caveat implying that you could be infected now even if you were not infected when the sample was taken)
New Zealand wording	New Zealand Ministry of Health	548	Indirectly, as a remote possibility ('We expect very few (if any) false positive test results (a false positive being a positive test result for someone who does not have the disease')	Yes (paragraph about reasons why false negatives are possible)
UK wording	UK National Health Service	443	No	No
Experimental wording	UK National Health Service (modified by researchers)	613	Indirectly, as a remote possibility ('A positive result means you almost certainly had coronavirus when the test was done')	Yes (paragraph about reasons why false negatives are possible)

*Wording retrieved from original sources in September 2020.

to a questionnaire in Qualtrics. Participants completed the questionnaire between 29 October and 2 November 2020. All participants gave informed consent before participation and were randomised afterwards using the 'randomise' function in Qualtrics. Participants were presented with the scenario 'John has been feeling ill. Based on his symptoms alone, a knowledgeable doctor believes that John has a 50–50 chance of having COVID-19. John takes the COVID-19 (swab) test to see whether he currently has the virus, and it is …' with the final word of the sentence being *positive* or *negative*, depending on whether they had been randomised into the *positive* or *negative* 'test result' condition. We also placed the final word of the sentence in bold font, to increase the likelihood that participants would read and remember the result. Participants were then further randomised to one of five 'information' conditions. Those in a control condition received no further information; others were given the explanatory text around interpreting COVID-19 test results used on official websites produced by one of the following: the UK's NHS, the USA's Centers for Disease Control, New Zealand's Ministry of Health or the experimental information described previously. All information that could be used by participants to identify the original source/country of the text was removed. The full texts given to participants are shown in the online supplemental information.

Participants were then asked a series of questions, including: 'What is your best guess as to the percent chance that John actually had COVID-19 at the time of his test, given his result?', which participants answered on a slider scale; questions about their impression of the accuracy, reliability, certainty and trustworthiness of the result, which were combined into an index measure of 'perceived trustworthiness'; the quality of the evidence underlying the test result; their confidence in their estimate of the chance that 'John' had COVID-19; and—for participants not in the control condition—how clear and easy to understand they felt the information to be, answered on 7-point Likert scales. The first three measures were part of our set of key dependent outcome variables as described earlier. Our index measure of perceived trustworthiness was based on the findings of the pilot, in which answers to

the questions on certainty, accuracy, reliability and trustworthiness were highly correlated (all $r > 0.7$, Cronbach's alpha=0.95), and were therefore averaged into a single index of perceived trustworthiness.

All participants were also asked 'Given John's test result, how much do you agree or disagree with the statement that 'John should now isolate himself from other people'?', which was our fourth key dependent variable. They answered on a slider scale between 'completely disagree' to 'completely agree', which was converted to 0–100 for analysis. Afterward, participants were asked 'What are your reasons for your answer above?' and given a free text response box. Free response answers were coded by one of the authors into categories capturing the most prominent themes that emerged from their assessment of the responses via inductive coding.[34] They were also asked a number of demographic questions and questions about their experience with, and perception of the risk of, COVID-19.

Timers were also set to measure the amount of time participants spent reading the information and answering the sets of questions. Full details of all questions and their answer options are shown in the online supplemental information.

In the pilot experiment, participants had been asked to type in their best guess as to the per cent chance that 'John' actually had COVID-19 at the time of his test. For the main experiment this was changed to a slider and labels added to confirm '0%: John definitely **did not** have COVID-19' and '100%: John definitely **did** have COVID-19'. This was done because in the pilot, close to half of participants typed in '50%' as their answer, which is well known as a response that some individuals use to indicate that they do not know the answer; this tendency can be mitigated by providing a slider.[17–19]

Power calculation

Based on the effect sizes achieved in the pilot, we calculated that 1643 participants would be required to achieve 90% power to detect an effect size of $\eta_p^2 = 0.009$ (f=0.097), and adjusted our total required sample upwards on the

assumption that the same proportion of participants would fail the attention check as did in the pilot. Consequently, we intended to sample 2041 participants; due to slight oversampling, our final sample contained 2049 participants.

Preregistered analyses

5 (message wording)×2 (test result) analyses of variance (ANOVAs) were preregistered for each primary dependent variable: participants' estimates of 'John's' chance of having COVID-19, the perceived quality of evidence behind 'John's' test result, the trustworthiness of 'John's' test result and level of agreement with the statement 'John should now isolate himself from other people'. One of the assumptions of an ANOVA is normality of residuals. Although the technique is reasonably robust to violations of this assumption,[35 36] violations can lead to increased probability of type I errors and loss of power,[36] so alternatives are sometimes suggested for severe normality violations. One reason to expect such violations is the aforementioned tendency of many participants to answer 50% to indicate 'I don't know' in studies where participants are asked to estimate probabilities of events,[17] such as for our first key dependent outcome measure which asked participants to estimate the chance (in per cent) that 'John' had COVID-19 at the time of the test. Although we implemented a slider scale to mitigate the problem, we still anticipated that a substantial proportion of participants would provide midpoint answers. We therefore noted in our preregistration that non-parametric tests would be substituted if ANOVA assumptions were sufficiently violated. We also aimed to look for differences among the different wording groups in the frequency of responses of ~50%, though this was not stated in the preregistration.

Because applying a simple rank transform prior to ANOVA can cause unreliable estimation of interactions, an approach known as aligned ranks transformation ANOVA[37] is recommended[38] for factorial designs. These models always included the *wording * test result* interaction term. For regular parametric ANOVAs, we ran the model with the interaction term (using type 3 sums-of-squares), but when there was no interaction, we excluded the interaction term and reported a main-effects-only model (type 2 sums-of-squares).

Patient and public involvement

Patients and the public were not involved in the design or coproduction of this study. The results will be disseminated to study participants who requested it.

RESULTS

Characteristics of the participants in the main study are shown in table 3. The questionnaire took a median of 14 minutes for participants to complete, and they were paid £0.75 for their participation. Analysis was conducted in R (V.3.6). Two hundred and ninety-seven participants did not pass the attention check and were excluded from all analyses. Final analytical sample sizes are reported in table 4.

Table 3 Participant characteristics

Participant characteristics		Number of participants (percentage)
Gender	Male	834 (48)
	Female	910 (52)
Age	18–24	194 (11)
	25–34	283 (16)
	35–44	319 (18)
	45–54	362 (21)
	55–64	312 (18)
	65+	274 (16)
Numeracy level*	1	64 (4)
	2	140 (8)
	3	300 (17)
	4	348 (20)
	5	338 (19)
	6	289 (17)
	7	97 (6)
	8	168 (10)
Education level	No formal education above age 16	225 (13)
	Professional or technical qualifications above age 16	325 (19)
	School education up to age 18	460 (26)
	Degree (Bachelors) or equivalent	528 (30)
	Degree (Masters) or other postgraduate qualification	185 (11)
	Doctorate	19 (1)

Participants quota-sampled from the UK, 2020.
*Numeracy measured by a combination of the adaptive Berlin numeracy test,[48] three questions from Schwartz et al,[49] and one from Lipkus et al.[50]

Residuals of the dependent variables of 'chance of COVID-19' (H1) and level of agreement with the statement that 'John' should self-isolate (H4) had fairly severe levels of non-normality, despite our use of slider scales. Aligned ranks transformation ANOVAs were therefore conducted for these two analyses.

Table 4 Number of participants in each condition

Message wording	Test result	
	Positive	Negative
Control	156	200
USA	165	177
UK	165	178
New Zealand	186	159
Experimental	200	158

Table 5 Main effects of test result

Research question	Mean (95% CI)		Median (95% CI)*		P value	η_p^2
	Positive test	**Negative test**	**Positive test**	**Negative test**		
Q1	72.96 (71.46 to 74.46)	38.38 (36.71 to 40.05)	75 (71 to 76)	50 (41.5 to 50)	<0.001	0.360
Q2	5.21 (5.12 to 5.30)	4.83 (4.74 to 4.92)	5.38 (5.26 to 5.38)	4.96 (4.78 to 5.02)	<0.001	0.021
Q3	5.17 (5.07 to 5.26)	4.75 (4.66 to 4.85)	5.35 (5.18 to 5.59)	4.92 (4.83 to 5.00)	<0.001	0.023
Q4	85.50 (84.12 to 86.88)	51.09 (48.85 to 53.33)	97 (93 to 98)	58.5 (54 to 61.5)	<0.001	0.287
Confidence in Q1 estimate (exploratory)	72.30 (70.66 to 73.94)	63.99 (62.21 to 65.77)	76.5 (74 to 79)	67 (62 to 68)	<0.001	0.027

*CIs of the median computed using the adjusted percentile bootstrap method.[51]

Q1: What influence did message wording and test result have on how likely participants thought it was that the test recipient actually had COVID-19?

An aligned ranks transformation ANOVA found a large main effect of test result ($F(1,1734)=976.2$, table 5), with those who were told that 'John's' test result was positive having a stronger belief that he had COVID-19 than those told his result was negative. However, there was no main effect of message wording and no interaction; this was also true after removing participants seemingly using 50% to indicate 'I don't know' (see online supplemental analysis 1). The results therefore did not confirm hypothesis H1. Exploratory one-way ANOVAs likewise found no main effect of message wording in either those who were presented with a positive test result or those presented with a negative result (figure 1).

Q2: What influence did message wording and test result have on participants' views of the quality of the evidence behind 'John's' test result?

We found a small main effect of test result on perceived quality of evidence ($F(1,1738)=37.09$, table 5): participants felt quality of evidence was higher for positive test results than negative test results (see also figure 2). As
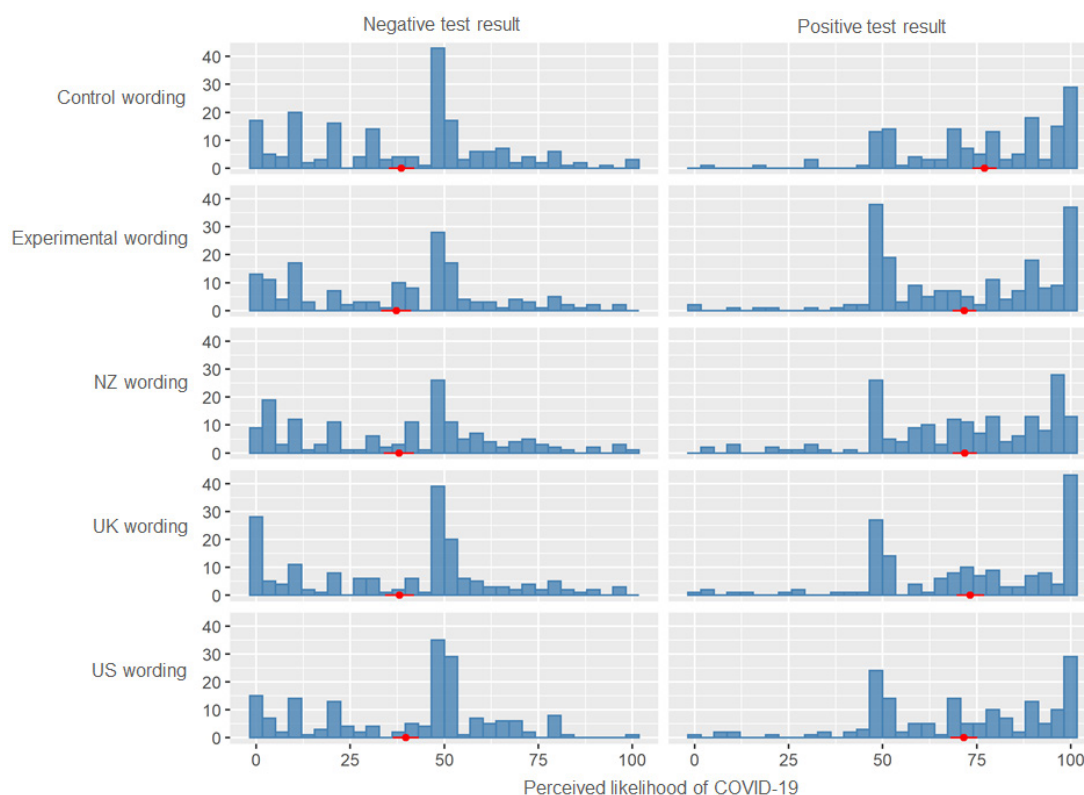

Figure 1 Distributions of participants' estimates of the chances of the test recipient 'John' actually having COVID-19, given either a negative (left hand side) or positive (right hand side) test result, and with either no accompanying explanatory text (control wording), an experimental text based on the UK wording but with added clarification about the test uncertainties (experimental wording), or official text from one of three different countries, that is, New Zealand (NZ wording), UK (UK wording), USA (US wording). The large hump around 50% is often interpreted as a response from participants who 'don't know'[17]. Red markers indicate group mean and 95% CI.

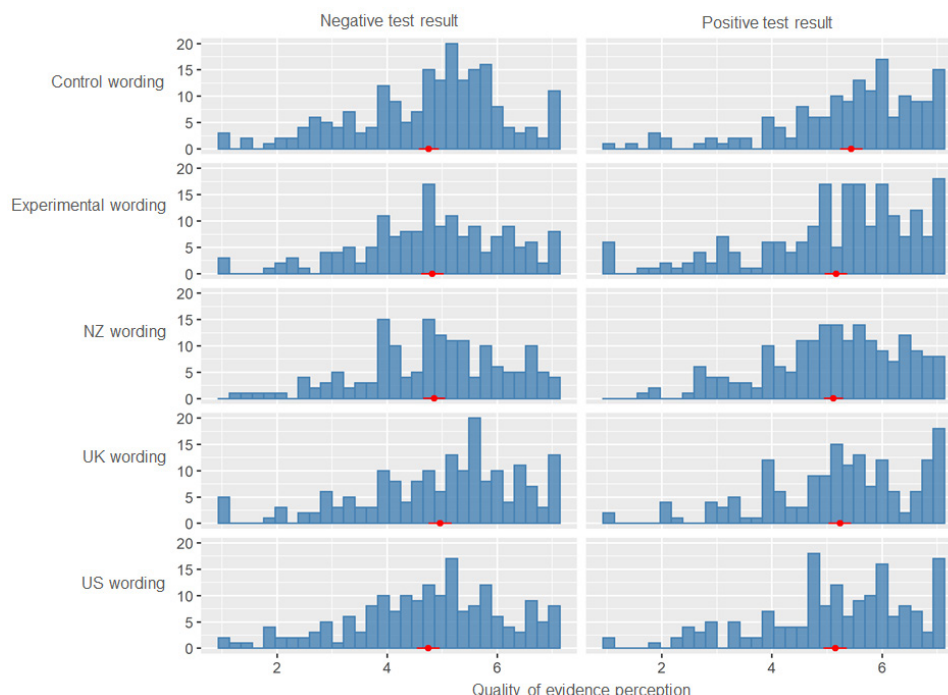Recchia G, *et al. BMJ Open* 2021;**11**:e047731. doi:10.1136/bmjopen-2020-047731

Figure 2 Distributions of participants' perception of the quality of the evidence underlying 'John's' PCR test result (rated on a 1–7 Likert scale with 1 indicating low quality of evidence and 7 indicating high quality of evidence) given either a negative (left hand side) or positive (right hand side) test result, and with either no accompanying explanatory text (control wording), an experimental text based on the UK wording but with added clarification about the test uncertainties (experimental wording), or official text from one of three different countries, that is, New Zealand (NZ wording), UK (UK wording), USA (US wording). Red markers indicate group mean and 95% CI.

we did not find a main effect of wording or a significant interaction, hypothesis H2 was not confirmed.

Q3: What influence did message wording and test result have on participants' views of the trustworthiness of 'John's' test result?

We found a small main effect of test result on perceived trustworthiness ($F_{(1,1738)}=40.42$, table 5), with participants perceiving the positive test results as more trustworthy than the negative test results (see also figure 3). We additionally found a small main effect of wording ($F_{(4,1738)}=2.41$, table 6), with Tukey's post-hocs suggesting that participants rated the test result as slightly less trustworthy if they read the NZ wording (mean=4.81, 95% CI 4.67 to 4.94) than if they read the UK wording (mean=5.10, 95% CI 4.96 to 5.25), p_{adj}=0.020. The effect of wording was only significant when the (non-significant) interaction term is removed from the model, and was not found in the pilot. As we did not find a significant interaction, hypothesis H3 was not confirmed.

Q4: What influence did message wording and test result have on participants' views about whether 'John' should self-isolate?

Aligned ranks transformation ANOVA revealed a large main effect of test result ($F_{(1,1734)}=699.3$, table 5): Most participants who were given a positive result for 'John' very much agreed with the statement that 'John' should isolate (median=97, 95% CI 93 to 98), whereas

participants who saw the negative condition exhibited more divergent opinions (median=58.5, 95% CI 54 to 61.5). Many of those who were told he had a negative result still agreed he should isolate: 41% (N=354) of those who were told he had a negative result still placed the slider at least two-thirds of the way towards the rightmost endpoint of the scale. This analysis also revealed a small main effect of wording ($F_{(4,1734)}=8.27$, table 6). CIs and post-hoc Mann-Whitney tests suggested more agreement with the statement among participants viewing the NZ wording than among those viewing the control or the UK wording (CIs in table 7). There was also a small interaction ($F_{(4,1734)}=8.51$, p<0.001, η_p^2=0.019), described later; means and CIs for each condition are given in table 7, and full distributions in figure 4.

The results partially confirm hypothesis H4, that there would be an interaction between wording and test result. Those shown a positive test result were more likely to agree that the recipient should self-isolate, but the differences among the wording groups were not as hypothesised. Instead, participants who viewed the NZ wording trended towards being more cautious in the case of a negative result (table 7). That said, 95% CIs between all pairs of wordings for a given test result overlapped at least slightly, and as mentioned, overall agreement that 'John' should self-isolate (aggregating across test results) was highest for those viewing the NZ wording.
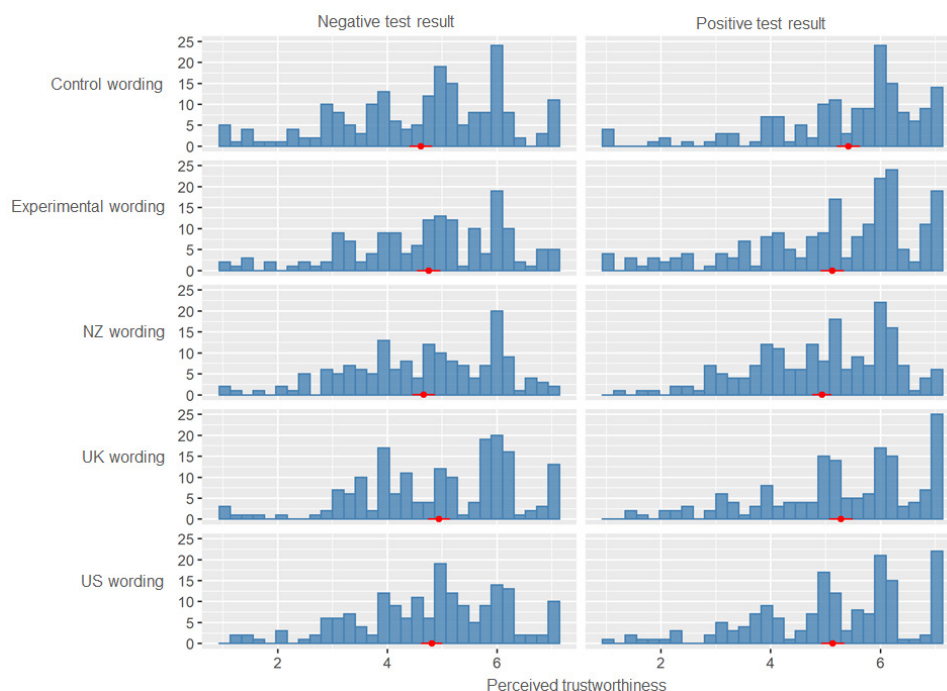
Figure 3 Distributions of participants' perception of the trustworthiness of 'John's' PCR test result (rated on a 1–7 Likert scale with 1 indicating low trustworthiness and 7 indicating high trustworthiness) given either a negative (left hand side) or positive (right hand side) test result, and with either no accompanying explanatory text (control wording), an experimental text based on the UK wording but with added clarification about the test uncertainties (experimental wording), or official text from one of three different countries, that is, New Zealand (NZ wording), UK (UK wording), USA (US wording). Red markers indicate group mean and 95% CI.

Exploratory analyses

50% responses: beliefs about 'John's' COVID-19 status

Overall for our measure of likelihood that 'John' had COVID-19 at the time of testing, 25.4% of participants gave an answer between 48% and 52% (a range we defined around the 50% midpoint given that people may not achieve exact accuracy with the slider), a phenomenon that is commonly observed and is attributed to '50–50' being akin to 'I don't know'.[17 18] There was no difference in the prevalence of this answer among the different wording groups with the possible exception of participants who saw the NZ wording, of whom only 19.1% gave answers between 48% and 52%: An exploratory χ^2 test suggested a possible difference (p<0.05) when participants in the NZ condition were included, but not when they were excluded. In addition to participants answering 50%, participants answering 0% or 100% were also of particular interest, as these answers arguably suggest more certainty than is warranted about the implications of a positive or negative test result.

Table 6	Main effects of wording						
Research question	**Mean (95% CI)**					**P value**	η_p^2
	Control	**USA**	**New Zealand**	**UK**	**Experimental**		
Q3	4.96 (4.81 to 5.12)	4.96 (4.82 to 5.11)	4.81 (4.67 to 4.94)	5.10 (4.96 to 5.25)	4.96 (4.81 to 5.11)	0.048	0.006
Q4	64.65 (60.95 to 68.35)	70.27 (66.93 to 73.61)	72.54 (69.41 to 75.68)	65.40 (61.71 to 69.09)	68.70 (65.36 to 72.04)	<0.001	0.019
Confidence in Q1 estimate (exploratory)	69.92 (67.31 to 72.52)	67.83 (64.99 to 70.67)	63.96 (61.18 to 66.73)	70.31 (67.51 to 73.12)	68.65 (65.97 to 71.33)	0.003	0.009
Perceived understanding (exploratory)	n/a	5.61 (5.48 to 5.75)	5.11 (4.96 to 5.27)	5.43 (5.28 to 5.59)	5.44 (5.30 to 5.58)	<0.001	0.017
Perceived effort required to understand (exploratory)	n/a	4.34 (4.16 to 4.53)	5.00 (4.84 to 5.17)	4.69 (4.49 to 4.89)	4.73 (4.54 to 4.91)	<0.001	0.018

Table 7 Mean agreement with the statement 'John should now isolate himself from other people' by condition

Message wording	Test result	
	Positive	**Negative**
	Mean (95% CI)	**Mean (95% CI)**
Control	88.07 (84.87 to 91.27)	46.38 (41.63 to 51.14)
USA	87.62 (84.93 to 90.31)	54.10 (49.21 to 58.98)
UK	83.27 (79.48 to 87.06)	48.84 (43.72 to 53.95)
New Zealand	85.44 (82.81 to 88.08)	57.46 (52.27 to 62.65)
Experimental	83.62 (80.55 to 86.70)	49.81 (44.62 to 55.00)

'Categorical' 100% or 0% responses: beliefs about 'John's' COVID-19 status

When asked how likely it was that 'John' actually had COVID-19 at the time of his test result, a χ^2 test suggested there were differences among wording groups in terms of likelihood to provide categorical '100%' or '0%' answers in the positive and negative test scenarios (100%: χ^2(4, N=872)=22.1, p<0.001; 0%: χ^2(4, N=872)=14.4, p=0.006; both: χ^2(4, N=1744)=32.0, p<0.001). Follow-up χ^2 tests and inspection of CIs suggested these differences were largely due to the participants who saw the UK wording being more likely to give categorical responses, and

Table 8 Participant responses to 'What is your best guess as to the percent chance that John actually had COVID-19 at the time of his test, given his test result?', by message viewed

Message viewed	N answering 48%–52% (percentage, 95% CI[52])	N answering categorically (0% or 100%) (percentage, 95% CI[52])
Control	84/356 23.6% (19.5% to 28.3%)	42/356 11.8% (8.8% to 15.6%)
USA	98/342 28.7% (24.1% to 33.7%)	42/342 12.3% (9.2% to 16.2%)
New Zealand	66/345 19.1% (15.3% to 23.6%)	20/345 5.8% (3.7% to 8.8%)
UK	97/343 28.3% (23.8% to 33.3%)	68/343 19.8% (15.9% to 24.4%)
Experimental	98/358 27.4% (23% to 32.2%)	41/358 11.5% (8.5% to 15.2%)

participants who saw the NZ wording less likely to (see table 8 (second column) and figure 5A,B). Further detail is available in table 9. χ^2 tests comparing the UK wording condition to each other group in turn suggested that those in the UK wording group were more likely to give a categorical response than every other group, and those in the NZ wording group were less likely to give a categorical
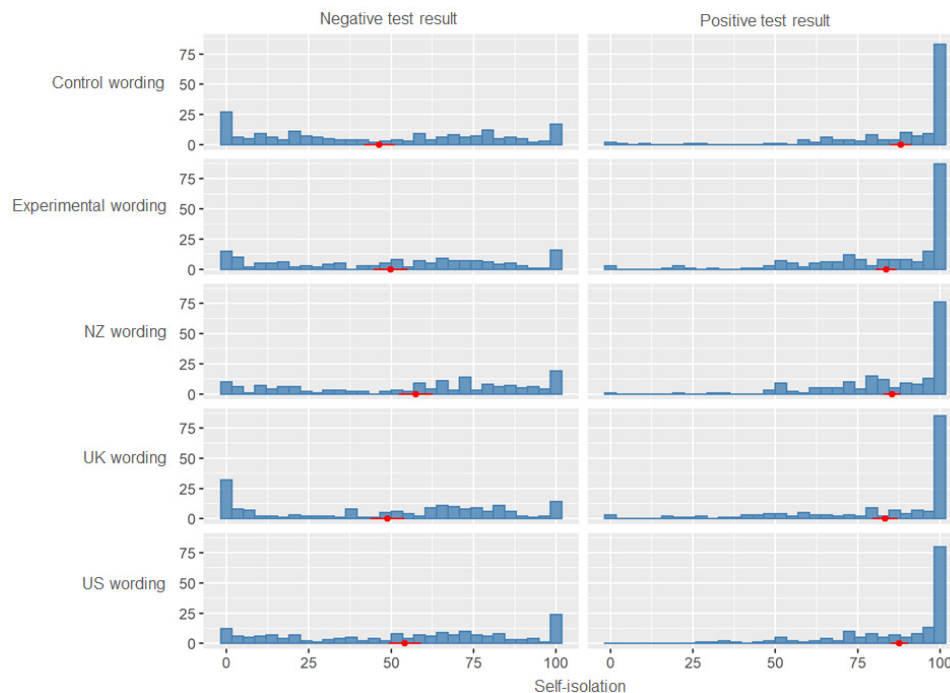

Figure 4 Distributions of participants' amount of disagreement (0) or agreement (100), with the statement that the test recipient 'John' should continue to self-isolate given either a negative (left hand side) or positive (right hand side) test result, and with either no accompanying explanatory text (control wording), an experimental text based on the UK wording but with added clarification about the test uncertainties (experimental wording), or official text from one of three different countries, that is, New Zealand (NZ wording), UK (UK wording), USA (US wording). Red markers indicate group mean and 95% CI.
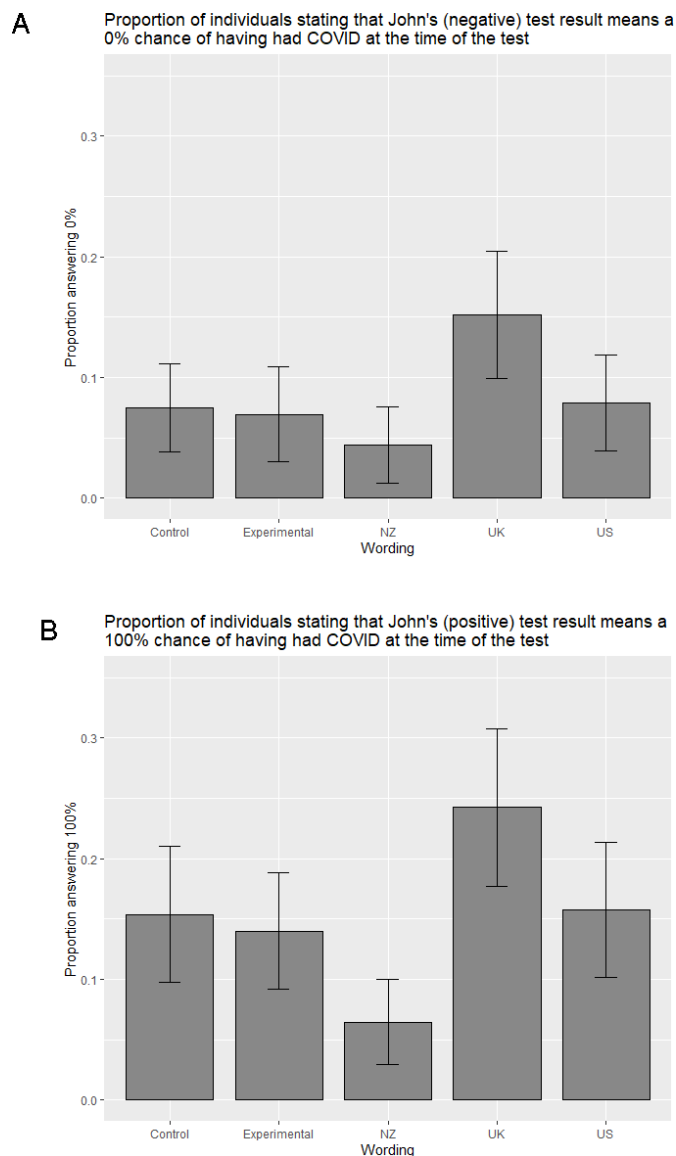
A

Proportion of individuals stating that John's (negative) test result means a 0% chance of having had COVID at the time of the test

B

Proportion of individuals stating that John's (positive) test result means a 100% chance of having had COVID at the time of the test

Figure 5 (A and B) The proportion of participants giving categorical 100% (A) or 0% (B) responses as to what 'John's' viral test result means in each message wording group, for those who were told that 'John's' result was (A) positive and (B) negative.

response than any other group, all tests p<0.05. (Whether p values lose their meaning in exploratory analyses is a matter of some debate. We report them for reference but acknowledge that their interpretation is controversial; see Rubin[39] for arguments on both sides.)

'Categorical' 100% or 0% responses: beliefs about whether 'John' should self-isolate

A χ^2 test on the subset of participants who were told that 'John' tested *negative* (after having been told he was experiencing symptoms) suggested that participants were more likely to express maximal disagreement with the statement 'John should now isolate himself from other people' if they were presented with the UK message (χ^2(4, N=872)=23.9, p<0.001): 17.4% of those in the UK wording group expressed maximal disagreement, in contrast to 3.8% (NZ), 5.1%

(USA), 8.9% (experimental) and 11.0% (control) (see table 10 for Ns and CIs). Further analyses were conducted on this subset of participants to see if these differences were likely to be meaningful: first, compared head-to-head with the control group, a χ^2 test suggested that the NZ wording group was less likely to express maximal disagreement, χ^2(1, N=359)=5.5, p=0.02. Second, head-to-head comparisons between the UK wording condition and each other group suggested that those in the UK wording group were more likely to express maximal disagreement than every other group except for the control, for which there was no clear difference. There was no clear evidence that participants who heard that 'John' had tested *positive* were more likely to express maximal agreement in one wording condition over another (χ^2(4, N=872)=5.3, p=0.256).

How easy or difficult did participants find the information to understand?

As might be expected, ratings of understanding of the different wordings were associated with the length and the complexity of the text (e.g., description of the statistical concepts of false negatives/false positives), with the NZ wording being rated the hardest to understand on average, and the US wording the easiest. A two-way ANOVA (test result+test wording) of ratings of 'how easy' people found the wording to understand and 'how completely' they felt they understood it (averaged into an index measure) suggested a small main effect of wording (F(3,1372)=7.89, table 6), with post-hoc Tukey tests implying that the NZ text was harder to understand than all three other wordings: UK (p_{adj}=0.012), USA (p_{adj}<0.001) and experimental wording (p_{adj}=0.009). An ANOVA of participants' ratings of how much effort they had to put into understanding the wording also suggested a small main effect of wording (F(3,1381)=8.53, table 6), with the US wording being lower on invested effort compared with all other groups, UK (p_{adj}=0.045), NZ (p_{adj}<0.001) and experimental wording (p_{adj}=0.014), in line with our pilot findings. Given the degree of skew in the understanding outcome measures, we also ran non-parametric analyses as a robustness check. Non-parametric findings were in line with the parametric findings.

How much confidence did participants have in their answer regarding their beliefs about 'John's' COVID-19 status?

An exploratory ANOVA found a small main effect of test result, F(1,1738)=48.6, with greater confidence for positive test results than negative (table 5). Likewise, there was a small main effect of wording, F(4,1738)=4.1, table 6; post-hoc Tukey tests suggested that participants seeing the NZ wording were less confident in their estimates of the likelihood that the test recipient had COVID-19 (mean=64.0, 95% CI 61.2 to 66.7) than those viewing the UK wording (mean=70.3, 95% CI 67.5 to 73.1), p=0.004, or the control (mean=69.9, 95% CI 67.3 to 72.5), p=0.004. As the confidence measure exhibited skew, we also ran non-parametric analyses as a robustness check. Results were largely in line with the parametric findings; both a

Table 9 Participant responses to 'What is your best guess as to the percent chance that John actually had COVID-19 at the time of his test, given his test result?', by test result and message viewed

Combination of test result and message viewed	N answering 48%–52% (percentage, 95% CI[52])	N answering 0% ('John definitely did not have COVID-19') (percentage, 95% CI[52])	N answering 100% ('John definitely did have COVID-19') (percentage, 95% CI[52])
Positive test			
Control	25/156 16.0% (11.0% to 22.6%)	0/156 0.0% (0% to 2.9%)	24/156 15.4% (10.5% to 21.9%)
USA	37/165 22.4% (16.7% to 29.4%)	1/165 0.6% (0% to 3.7%)	26/165 15.8% (10.9% to 22.1%)
New Zealand	30/186 16.1% (11.5% to 22.1%)	0/186 0.0% (0% to 2.4%)	12/186 6.5% (3.6% to 11.0%)
UK	39/165 23.6% (17.8% to 30.7%)	1/165 0.6% (0% to 3.7%)	40/165 24.2% (18.3% to 31.3%)
Experimental	57/200 28.5% (22.7% to 35.1%)	2/200 1.0% (0% to 3.8%)	28/200 14.0% (9.8% to 19.5%)
Negative test			
Control	59/200 29.5% (23.6% to 36.2%)	15/200 7.5% (4.5% to 12.1%)	3/200 1.5% (0.3% to 4.5%)
USA	61/177 34.5% (27.9% to 41.7%)	14/177 7.9% (4.7% to 12.9%)	1/177 0.6% (0% to 3.5%)
New Zealand	36/159 22.6% (16.8% to 29.8%)	7/159 4.4% (2.0% to 9.0%)	1/159 0.6% (0% to 3.8%)
UK	58/178 32.6% (26.1% to 39.8%)	27/178 15.2% (10.6% to 21.2%)	0/178 0% (0% to 2.5%)
Experimental	41/158 25.9% (19.7% to 33.3%)	11/158 7.0% (3.8% to 12.2%)	0/158 0% (0% to 2.9%)

main effect of test results and a main effect of wording emerged. Uncorrected Mann-Whitney tests additionally suggested a difference between the NZ wording and the US wording as well as the experimental wording. See figure 6.

Free text responses

Free text answers to why participants felt the test recipient should self-isolate were thematically coded by one of the authors into 1 of 10 themes: mistakenly remembering the direction of the test result; scepticism about the seriousness or existence of COVID-19; uncertainty about the validity of the test result (including quality of evidence concerns); John's test result means he does not need to self-isolate; concern for the mental health, personal freedom, practical or financial implications of self-isolation; scepticism about the effectiveness of self-isolation; John's symptoms mean he should self-isolate; John's test result means he should self-isolate; general comments about self-isolation being important for public health, or 'better safe than sorry'; other/not sure/forgotten result. Where more than one reason was given, the response was scored on the first mentioned (see figure 7).

For exploratory analyses looking at the effects of numeracy, education and ethnicity see online supplemental analysis 2.

DISCUSSION

Test, trace and isolate systems (both population-wide and as a specific requirement for entering work or educational environments) have been a core part of many countries' responses to COVID-19, and they rely greatly on people's behaviour upon receiving either a positive or negative test result. Studies have looked at the public's behaviour in the real world when experiencing symptoms,[40] or at factors affecting adherence to quarantine guidelines, which is generally low,[41–43] although of course not just a matter of intentions, but also of difficulty.[44 45] To our knowledge no studies have looked at the effects of information on, and interpretation of, COVID-19 test results as a part of this important pandemic mitigation response.

The interpretation of a positive or negative test result is not simple, because it depends on the prior likelihood of the condition. As of 2020, the UK's NHS had decided to communicate the results of COVID-19 PCR tests with no uncertainty, in an apparent attempt to maximise clarity for the recipient; changes introduced in 2021 explicitly acknowledged a degree of uncertainty. New Zealand's Ministry of Health had, by contrast, dedicated an entire website to trying to communicate the uncertainties. The USA's Centers of Disease Control had taken the middle ground, merely adding the caveat that a recipient of a

Table 10 Participant responses to 'John should now isolate himself from other people', by test result and message viewed

Combination of test result and message viewed	N with maximum disagreement (slider all of the way to 'completely disagree') (percentage, 95% CI[52])	N with maximum agreement (slider all of the way to 'completely agree') (percentage, 95% CI[52])
Positive test		
Control	1/156 0.6% (0% to 3.9%)	78/156 50.0% (42.2% to 57.8%)
USA	0/165 0% (0% to 2.7%)	75/165 45.5% (38.0% to 53.1%)
New Zealand	1/186 0.5% (0% to 3.3%)	75/186 40.3% (33.5% to 47.5%)
UK	2/165 1.2% (0.1% to 4.6%)	80/165 48.5% (41% to 56.1%)
Experimental	3/200 1.5% (0.3% to 4.5%)	82/200 41.0% (34.4% to 47.9%)
Negative test		
Control	22/200 11.0% (7.3% to 16.2%)	16/200 8.0% (4.9% to 12.7%)
USA	9/177 5.1% (2.6% to 9.5%)	22/177 12.4% (8.3% to 18.2%)
New Zealand	6/159 3.8% (1.6% to 8.2%)	18/159 11.3% (7.2% to 17.3%)
UK	31/178 17.4% (12.5% to 23.7%)	12/178 6.7% (3.8% to 11.5%)
Experimental	14/158 8.9% (5.2% to 14.4%)	15/158 9.5% (5.7% to 15.2%)

Figure 6 Participant confidence in their own estimates of the likelihood that the test recipient had COVID-19, by wording group (either no accompanying explanatory text (control wording), an experimental text based on the UK wording but with added clarification about the test uncertainties (experimental wording), or official text from one of three different countries, that is, New Zealand (NZ wording), UK (UK wording), USA (US wording)).

and assessed it as more trustworthy. Exploratory analyses also implied that participants expressed more confidence in positive than negative results.

This suggests that, even with no information provided (i.e., in the control group), the UK public are aware that such test results are not perfectly accurate, and have a lower confidence in a negative test result than a positive test result. This is a reasonable belief for the scenario presented—where the pretest probability of having COVID-19 was stated as 50%—given that evidence from clinical adjudication, systematic reviews and community-based settings suggests that specificity has been higher than sensitivity for PCR tests in real-world settings[1].

Differences in response to the different information (or none) accompanying results

Some differences in response were also seen among the groups that received different wording alongside the test result. In preregistered analyses, there was a suggestion that those seeing the NZ wording found the results slightly less trustworthy, and they were also more likely than those who saw the UK wording to say that the test recipient should self-isolate. Exploratory findings built on these, suggesting that those who read the UK wording's unambiguous interpretation of the results tended to be more definitive in their interpretation as well: they were more likely to answer '100%' as the likelihood of the recipient of a positive test result having COVID-19 and '0%' as the likelihood of the recipient of a negative test result having the disease. By contrast, those who read the NZ wording were the least likely to be so categorical in their interpretations (as well as being the least likely to answer '50%', which can be taken as an indication of 'I don't know'). For those who had been shown negative test results, this trend carried through to participants'

negative test result 'probably' does not have COVID-19. These three pieces of information are not comparable in their aims and audience—they are not all provided to test recipients when they are given their results—but we here compare their effects on the public's interpretation of test results as a way of gauging how such information affects people's natural understanding of such a test result.

Differences between response to positive versus negative test results

Preregistered analyses showed that participants thought the (symptomatic) recipient of a positive test result was more likely to have COVID-19 than the (symptomatic) recipient of a negative test result, and were more likely to say that such a recipient should then self-isolate. They also rated the positive test result to have a higher quality of evidence underlying it,

Figure 7 The reasons participants gave for their response to how much they agreed with the statement that the test recipient (who was symptomatic) should isolate after receiving his test result, split by the result they were given for the recipient (positive or negative) and the additional information that they received alongside the test result (either no accompanying explanatory text (control wording), an experimental text based on the UK wording but with added clarification about the test uncertainties (experimental wording), or official text from one of three different countries, that is, New Zealand (NZ wording), UK (UK wording), USA (US wording)).

reactions to whether or not the recipient should self-isolate or not.

The fact that the effect of the UK's official interpretation of the result appeared to lead some participants to interpret a positive result as meaning that the recipient was 100% likely to have COVID-19 is not a major concern; 100% may not be far off the true rate given the low false positive rate for symptomatic individuals. However, what is more worrying is that, as some commentators had feared,[7] the wording also appeared to lead some participants to believe that a negative result meant the recipient had a 0% chance of having COVID-19, and that this appeared to carry through to their beliefs that he should not self-isolate, despite being symptomatic. The experimental information, based on the UK wording but incorporating what we hoped was clearer advice around a negative test result, did not appear to discourage this view as much as we had hoped, although the free text responses in this group did seem to suggest an increase in the number of people who picked up on the fact that a symptomatic person should continue to self-isolate.

Those in the NZ wording condition also, in exploratory findings, rated themselves less confident in their judgments of the likelihood that the test recipient had COVID-19 at the time of the test, given the test results. The NZ wording, with its emphasis on the uncertainty of the tests, appears, then, to have influenced participants in this wording condition to come away with a more nuanced view of what a COVID-19 test result means: they thought the results less trustworthy, were less likely to think them definitive, and were more cautious in their behavioural interpretation of a negative test. Whether their lower confidence in their answer is a good or bad thing is arguable.

It was also interesting that those seeing the UK wording seemed to interpret results not only as more definitive than those in the other wording conditions, but also than those in the control group (table 8), which yields insight into people's beliefs about test result interpretation prior to being shown further information.

Finally, exploratory results suggested that participants rated the NZ wording harder to understand, and the US wording as the one requiring the least effort, both in line with our manipulation check of time spent reading the intervention texts (see online supplemental information). The US wording was shorter than the NZ wording, so it would be expected that participants take more time to read the NZ wording. Additionally, given that the NZ text provided information on statistical uncertainty (e.g., false positives and false negatives), which requires more cognitive effort in order to comprehend the content compared with a text that does not describe uncertainties in such depth, the findings of understanding and effort invested are not too surprising.

The free text that participants gave to explain why they thought the test recipient should or should not self-isolate often emphasised a precautionary approach. A large number chose to be 'better safe than sorry', although

fewer among those who were given information about test interpretation from the UK's NHS website. From those who saw the NZ wording and a negative test result, the free text responses suggest that the larger proportion of people saying he should self-isolate were doing so because of uncertainties about the test result (rather than because he was symptomatic). Very few participants were sceptical about the seriousness or existence of COVID-19, about the effectiveness of self-isolation, or that the potential negative consequences of self-isolation (e.g., financial, effects on personal freedom or mental health) outweighed any potential positives in terms of public health. These free text responses bolster the interpretation of the quantitative results.

This study has a number of limitations. The scenario given to participants, of a high suspicion (50% chance) that the test recipient had COVID-19 was not clinically likely unless they were a health worker (or already hospitalised because of symptoms) given the prevalence of COVID-19 at the time our study was conducted. We did not introduce either of these factors as we felt they would be likely to bias participants' interpretation of either the test result or their behavioural recommendation. However, depending on overall prevalence, in many cases the prior probability will be much lower than 50% and we have not tested how this would affect participants' interpretations. We also presented participants with text taken from official websites, rather than the much shorter text they are likely to receive (e.g., by SMS message) alongside their test result. Additionally, the question asking participants to give a numerical probability of the chances of the likelihood of COVID-19 still resulted in a high proportion of people indicating that they were uncertain how to respond.

There were also differences in results between the pilot and the main study. Some of these may be due to improvements we made to the questionnaire after the pilot: the difference in how participants were asked to input percentages when asked the test recipient's chance of having COVID-19 (a slider vs a text box in the pilot), which decreased the number of participants who provided answers of 50%—although many people still answered 50% even when using the sliderscale—and putting the test result in bold in an effort to ensure that it was not missed.

Finally, our study participants were only UK residents who responded to the online survey invitation; this means they should not necessarily be taken as representative of the UK public as a whole, and the public of other countries may have had different reactions.

Our conclusions, however, are that although the UK public have a natural sense of the comparative reliability of positive and negative test results when the prior probability of infection is high, presenting them with an unambiguous interpretation encourages a belief that the results are 100% certain.

CONCLUSIONS

Test results are often misinterpreted, particularly if the prevalence of the condition being tested for is very low.[46]

The overwhelmingly large effect of the prior probability of a positive test (strongly influenced by the local prevalence, as well as by the presence and type of symptoms) on the interpretation of a test result is not only unintuitive, it also makes it difficult to communicate the numerical uncertainty around an individual's test result, because this prior probability is so variable. Added to these complications are the difficulties associated with assessing sensitivity and specificity of COVID-19 PCR tests in a real-world context, where additional uncertainties are introduced by users carrying out the swabbing themselves, as opposed to a laboratory context. However, the results of this study show that the public are quite capable of understanding more nuanced explanations than many authorities give them credit for.

It may also be worth communicating that negative results should be confirmed with a second test when symptomatic, although even this should not be communicated in such a way as to give undue confidence. With plausible assumptions, Watson *et al*'s[1] table 1 estimates that someone with a prior probability of 50% will have a post-test probability of 24% after a single negative test, and of 9% after two negative tests assuming independence of test failures, though in the real world causes of test failure can be correlated.

We applaud the UK NHS for modifying their original 2020 website text in such a way as to acknowledge uncertainty and to be more clear about self-isolation, in line with our suggestions. This seems likely to discourage the belief that results are 100% certain: our experimental wording was directly based on the UK wording with a few small tweaks, and did not appear to inspire the same degree of categorical interpretation. However, it did not appear to affect the belief that a symptomatic person with a negative test result could stop self-isolating. Only the NZ wording had this effect, and other agencies might consider adopting wording more similar to that used by the New Zealand Ministry of Health. Its rating of being harder to understand (and possible follow-on effects to the understanding of a positive test) should be borne in mind, though we did not observe evidence that those reading this wording would be less apt to self-isolate after a positive test. Importantly, the straightforward advice that a symptomatic person should continue to self-isolate after a single negative test result needs to be given very strongly in all governments' information, in line with UK NHS and government guidance which states that individuals who feel unwell should continue to stay at home.[3 47]

ORCID iDs
Gabriel Recchia http://orcid.org/0000-0002-0210-8635
Claudia R Schneider http://orcid.org/0000-0002-6612-5186
Alexandra LJ Freeman http://orcid.org/0000-0002-4115-161X

REFERENCES

1 Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. *BMJ* 2020;369:1–7.
2 Office for National Statistics (ONS). Coronavirus (COVID-19) infection survey, UK: 20 November 2020 [online], 2020. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/20november2020
3 NHS. Your coronavirus test result, 2020. Available: https://www.nhs.uk/conditions/coronavirus-covid-19/testing-and-tracing/what-your-test-result-means/ [Accessed September 2020].
4 Centers for Disease Control. What your test results mean, 2020. Available: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html [Accessed September 2020].
5 Centers for Disease Control. Test for current infection, 2020. Available: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html [Accessed September 2020].
6 Ministry of Health NZ. COVID-19 test results and their accuracy, 2020. Available: https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-health-advice-public/assessment-and-testing-covid-19/covid-19-test-results-and-their-accuracy [Accessed September 2020].
7 McCartney M, Sullivan F, Heneghan C. Information and rational decision-making: explanations to patients and citizens about personal risk of COVID-19. *BMJ Evid Based Med* 2020.
8 Black WC, Nease RF, Tosteson AN. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *J Natl Cancer Inst* 1995;87:720–31.
9 Hurd MD, McGarry K. The predictive validity of subjective probabilities of survival. *Econ J* 2002;112:966–85.
10 Bruine de Bruin W, Downs JS, Murray P, *et al*. Can female adolescents tell whether they will test positive for Chlamydia infection? *Med Decis Mak* 2010;30:189–93.
11 Persoskie A. How well can adolescents really judge risk? Simple, self reported risk factors out-predict teens' self estimates of personal risk. *Judgm Decis Mak* 2013;8:1–6.

12. Weinstein ND, Marcus SE, Moser RP. Smokers' unrealistic optimism about their risk. *Tob Control* 2005;14:55–9.
13. Weinstein ND, Kwitel A, McCaul KD, *et al*. Risk perceptions: assessment and relationship to influenza vaccination. *Health Psychol* 2007;26:146–51.
14. Weinstein ND. What does it mean to understand a risk? Evaluating risk comprehension. *J Natl Cancer Inst Monogr* 1999;25:15–20.
15. Fishbein M, Raven BH. The AB scales: an operational definition of belief and attitude. *Hum Relations* 1962;15:35–44.
16. Hurd MD, McGarry K. Evaluation of the subjective probabilities of survival in the health and retirement study. *J Hum Resour* 1995;30:S268–92.
17. Fischhoff B, Bruine De Bruin W. Fifty-Fifty=50%? *J Behav Decis Mak* 1999;12:149–63.
18. Fischhoff B, Millstein SG, *et al*. Verbal and numerical expressions of probability: "it's a fifty-fifty chance". *Organ Behav Hum Decis Process* 2000;81:115–31.
19. Bruine de Bruin W, Carman KG. Measuring subjective probabilities: the effect of response mode on the use of focal responses, validity, and Respondents' evaluations. *Risk Anal* 2018;38:2128–43.
20. Gustafson A, Rice RE. A review of the effects of uncertainty in public science communication. *Public Underst Sci* 2020;29:614–33.
21. van der Bles AM, van der Linden S, Freeman ALJ, *et al*. The effects of communicating uncertainty on public trust in facts and numbers. *Proc Natl Acad Sci U S A* 2020;117:7672–83.
22. Retzbach A, Maier M. Communicating scientific uncertainty: media effects on public engagement with science. *Communic Res* 2015;42:429–56.
23. Howe LC, MacInnis B, Krosnick JA, *et al*. Acknowledging uncertainty impacts public acceptance of climate scientists' predictions. *Nat Clim Chang* 2019;9:863–7.
24. Jensen JD. Scientific uncertainty in news coverage of cancer research: effects of hedging on scientists and Journalists credibility. *Hum Commun Res* 2008;34:347–69.
25. Maxim L, Mansier P, Maxim L. How is scientific credibility affected by communicating uncertainty? The case of endocrine disrupter effects on male fertility risk perception/communication articles how is scientific credibility affected by communicating uncertainty? *The Case* 2014;7039.
26. Nakayachi K, Johnson BB, Koketsu K. Effects of acknowledging uncertainty about earthquake risk estimates on San Francisco bay area residents' beliefs, attitudes, and intentions. *Risk Analysis* 2018;38:666–79.
27. Rosenbaum L. Communicating uncertainty — Ebola, public health, and the scientific process. *N Engl J Med Overseas Ed* 2015;372:7–9. doi:10.1056/NEJMp1413816
28. Johnson BB, Slovic P. Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. *Risk Anal* 1995;15:485–94.
29. Broomell SB, Kane PB. Public perception and communication of scientific uncertainty. *J Exp Psychol Gen* 2017;146:286–304.
30. Fischhoff B, Davis AL. Communicating scientific uncertainty. *Proc Natl Acad Sci U S A* 2014;111:13664–71.
31. Han PKJ. Conceptual, methodological, and ethical problems in communicating uncertainty in clinical evidence. *Med Care Res Rev* 2013;70:14S–-36S.
32. Wegwarth O, Wagner GG, Spies C, *et al*. Assessment of German public attitudes toward health communications with varying degrees of scientific uncertainty regarding COVID-19. *JAMA Netw Open* 2020;3:e2032335.
33. Blastland M, Freeman ALJ, van der Linden S, *et al*. Five rules for evidence communication. *Nature* 2020;587:362–4.
34. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. *Am J Eval* 2006;27:237–46.
35. Glass GV, Peckham PD, Sanders JR. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Rev Educ Res* 1972;42:237–88.
36. Cochran WG. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics* 1947;3:22.
37. Wobbrock JO, Findlater L, Gergle D. The aligned RANK transform for nonparametric factorial analyses using only ANOVA procedures. in: SIGCHI conference on human factors in computing systems 2011:143–6.
38. Nussbaum EM. *Categorical and nonparametric data analysis: choosing the best statistical technique*. Routledge, 2014.
39. Rubin M. Do p values lose their meaning in exploratory analyses? It depends how you define the familywise error rate. *Rev Gen Psychol* 2017;21:269–75.
40. Smith LE, Potts HWW, Amlôt R, *et al*. Adherence to the test, trace and isolate system: Results from a time series of 21 nationally representative surveys in the UK (the COVID-19 Rapid Survey of Adherence to Interventions and Responses [CORSAIR] study). *medRxiv* 2020:1–47.
41. Carlucci L, D'Ambrosio I, Balsamo M. Demographic and attitudinal factors of adherence to quarantine guidelines during COVID-19: the Italian model. *Front Psychol* 2020;11:559288.
42. Vagnoni C. Test, trace and isolate: behavioural aspects [online]. London, 2020. Available: https://post.parliament.uk/test-trace-and-isolate-behavioural-aspects/
43. Steens A, Freiesleben de Blasio B, Veneti L, *et al*. Poor self-reported adherence to COVID-19-related quarantine/isolation requests, Norway, April to July 2020. *Euro Surveill* 2020;25:1–6.
44. Reicher S, Drury J. Pandemic fatigue? How adherence to covid-19 regulations has been misrepresented and why it matters. *BMJ* 2021;372:1–2.
45. Cevik M, Baral SD, Crozier A, *et al*. Support for self-isolation is critical in covid-19 response government action needed now to reduce infections and deaths 2021:10–11.
46. Bramwell R, West H, Salmon P. Health professionals' and service users' interpretation of screening test results: experimental study. *BMJ* 2006;333:284–6.
47. Public Health England. Stay at home: guidance for households with possible or confirmed coronavirus (COVID-19) infection [online], 2020. Available: https://www.gov.uk/government/publications/covid-19-stay-at-home-guidance/stay-at-home-guidance-for-households-with-possible-coronavirus-covid-19-infection
48. Cokely ET, Galesic M, Schulz E, *et al*. Measuring risk literacy: the Berlin Numeracy test. *Judgm Decis Mak* 2012;7:25–47.
49. Schwartz LM, Woloshin S, Black WC, *et al*. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med* 1997;127:966–72.
50. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Mak* 2001;21:37–44.
51. Davison AC, Hinkley D. Confidence intervals. In: *Bootstrap methods and their application*. Cambridge: Cambridge University Press, 1997: 191–255.
52. Agresti A, Coull BA. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat* 1998;52:119–26.