Supplementary Materials: Broad Cross-National Public Support for Accelerated COVID-19 Vaccine Trials

Contents

Additional Detail on Survey

Survey Procedures

Before conducting the surveys, we filed a pre-registration and pre-analysis plan at https://osf.io/cgbxa/, the full text of which is available at the end of the SM. We describe two minor departures from our pre-registered pre-analysis plan in the last section.

We then conducted national surveys in Australia, Canada, Hong Kong, New Zealand, South Africa, Singapore, the United Kingdom, and the United States. We recruited participants online using the sample provider Lucid in early May, 2020. Survey respondents in Australia, the United Kingdom, and the United States surveys were selected to match Census benchmarks. We requested 2,000 participants in the United States and 500 for each other geography. Respondents must take the survey on a desktop computer and be able to read English.

In the survey, we first asked for consent to participate. 471 participants did not consent and were removed from the survey. 164 participants failed a first attention check, which asked participants to select "I understand" in response to "For our research, careful attention to survey questions is critical!"; these participants were removed. As a final attention check, we wrote: "People are very busy these days and many do not have time to follow what goes on in the government. We are testing whether people read questions. To show that you've read this much, answer both 'extremely interested' and 'very interested.'" We removed 1,636 participants for not reading carefully enough to select both 'extremely interested' and 'very interested'. 895 did not complete the survey. 5,920 participants remained and completed the survey.

We then told participants:

> COVID-19, often called coronavirus, has led to hundreds of thousands of deaths and millions of job losses.

> Many experts say that people will continue dying of coronavirus and daily life will not return to normal until there is a vaccine. A vaccine would make many people immune to coronavirus, so

they could not catch coronavirus or give it to other people.

Politicians and experts are debating several approaches to developing a coronavirus vaccine. These approaches are all medically and scientifically valid, but have different strengths and weaknesses. Because the coronavirus crisis affects everyone, many politicians and experts want to know what people think about the best approach. We would like to hear your opinion about which of two approaches you think the government and researchers should take. We plan to share our findings, so please take this survey seriously.

Do you understand?

We then gave the following background on vaccines:

To test whether a new coronavirus vaccine is safe and effective, scientists will accept volunteers for a study on vaccines. Scientists then will use a device (like flipping a coin) that randomly assigns volunteers in the study to either receive the vaccine or a placebo, a safe substance like salt water that does not have any vaccine in it. The scientists will then monitor whether study participants who received the vaccine instead of the placebo are less likely to get the coronavirus. Scientists will know the vaccine works if the participants who received the vaccine were less likely to get coronavirus than the people who received the placebo.

Do you understand?

We then asked two questions on people's general views on vaccines: "In your opinion, how important is it that parents get their children vaccinated?" and "Do you think vaccines are more dangerous than the diseases they are designed to prevent, or not?"

We then randomized participants to one of two studies: Study 1, about Challenge Trials, and Study 2, about Integrated Trials. We always began:

We would first like to get your opinion about two hypothetical ways to do a study to determine whether a new coronavirus vaccine is effective.

- There have already been initial studies that gave the vaccine to a small number of volunteers. In these initial studies, the vaccine appeared safe. [*For challenge trial only:* These initial studies also found that the vaccine has a good chance of working, even if it has not been proven effective yet.]

- For both of the studies we will describe, people have already volunteered to participate. Both studies are practical to start right away.

- People who volunteered to participate in the studies were fully informed about what each study involved before they volunteered.

- If any participants get sick during the studies, their medical care and any lost wages would be paid for. They would also have access to remdesivir, a drug that can often, but not always, improve time to recovery.

We described both studies as practical to begin to ensure respondents did not select on the basis of perceived practicability, as the ethical question is only of interest if both trials are practical to conduct.

We show how we described the trial designs to participants and the outcome variables in the context of the studies below. We did not allow participants to move on from the page describing the trial design until at least 60 seconds had gone by.

We finished the survey with a series of demographic questions regarding age, gender, education, political ideology, employment, religiosity, scientific knowledge, race/ethnicity, and, in the United States only, political party identification and zip code.

Demographics

Survey participants reported the demographic characteristics shown in Table S1. As expected for an online sample, the survey participants are slightly more educated than the population at large on average. Later we show the results are consistent for participants without a college degree. Survey respondents in Australia, the United Kingdom, and the United States surveys were selected to match Census benchmarks on age, gender, and race and ethnicity, but this was not possible in the other geographies. Finally, for the US sample, we later present an analysis weighting our sample to be representative of the US population; the findings do not change. Sample sizes and demographics shown are for the entire sample across both studies; respondents were randomly assigned to the two studies with equal probabilitiy, so are approximately evenly split across the two studies.

Table S1: Survey Demographics by Geography

	AUS	CAN	HK	NZ	SA	SG	UK	US	All
Average Age	49	47	33	44	36	33	43	49	44
% Female	52	52	48	45	46	47	47	63	54
% College Graduate	43	46	64	39	35	61	47	51	49
% Non-White (In HK & SG, Non-Chinese)	20	27	25	27	53	21	16	21	25
% Employed as an Essential Worker	24	25	67	22	27	43	22	27	30
% Employed as a Non-Essential Worker	25	26	15	28	24	28	31	19	23
% Unemployed due to COVID	5	4	1	4	10	2	3	6	5
% Furloughed due to COVID	4	7	3	8	12	5	12	5	7
N	500	687	422	498	548	520	565	2,180	5,920

For the US sample, relevant questions also closely match national averages from other surveys:

- Gallup data from December 2019 finds that 86% of Americans say that vaccines are not "more dangerous than the diseases they are designed to prevent"; in our US sample, this number is 83%. (In our sample outside the US, this figure is 80%.)
- Pew data from September 2014 finds that 76% of Americans know that ocean tides are caused by the pull of the moon and that 72% of Americans know that cell phones use radio waves; in our US sample, these figures are 77% and 73%, respectively. (In our sample outside the US, these figures are 73% and 75%.)

Full Survey Questions and Results

Study 1: Challenge Trial

Vaccine Study Design Description

After the preamble quoted above, participants were shown a table that looks like the example given in Figure S1. All of the highlighted elements were randomized. We did not allow participants to move on from the page describing the trial design until at least 60 seconds had gone by.

The example in Figure S1 shows the results of one particular randomization. All the highlighted numbers in the example were randomized, as detailed below. (These highlights did not appear for respondents.) In the below, the bolded numbers correspond with the numbers used in the example in Figure S1. We randomized these parameters given uncertainty about how particular vaccine trials might be conducted, to ensure our findings were not sensitive to any of these parameters.

- Standard Design Trial N: (3,000; 5,000; 7,000; **9,000**; 11,000)
    - N in each condition is calculated as half of this number.

Name	Study A	Study B	Key Differences
Study procedures	1. Recruit 80 healthy volunteers aged 18-30 to participate. 2. Give 40 people the vaccine and another 40 people a placebo. 3. Scientists expose all 80 participants to the coronavirus. Scientists keep these participants in a medical research center where they cannot infect others, are closely monitored, and provided any necessary medical care. 4. Scientists monitor participants for several weeks to learn if the vaccine works.	1. Recruit 9,000 healthy volunteers to participate. 2. Give 4,500 people the vaccine and another 4,500 people a placebo. 3. All participants then go about their daily lives. Scientists monitor participants for several months to learn if the vaccine works.	• In Study B, people catch the coronavirus on their own as they go about their daily lives. In Study A, participants are intentionally exposed to the virus.
Risks to study participants	• 40 participants in the placebo group are intentionally exposed to coronavirus while quarantined in a medical research center. Because all participants in this study are young people, it is unlikely any would develop serious complications. In addition, participants would likely develop immunity, so are unlikely to get coronavirus again or pass it along to others. • 40 participants who received the vaccine are also intentionally exposed to coronavirus. If the vaccine in the study does not work, they would also get coronavirus. In the worst-case scenario, they could get an especially bad case of it. • It is very unlikely any study participants would die of coronavirus.	• 90 participants in the placebo group are expected to get coronavirus from their daily lives. • It is likely 1 participant would die of coronavirus. • If the vaccine in the study does not work, 90 participants who received the vaccine would also get coronavirus. In the worst-case scenario, they could get an especially bad case of it.	If the vaccine works: • 50 more volunteers catch the coronavirus in Study B. • 1 more volunteer would die in Study B.
Expected time until vaccine ready	If the vaccine works, it would start being distributed widely in 6 months from today, November 2020.	If the vaccine works, it would start being distributed widely in 12 months from today, May 2021.	• If the vaccine works, Study A would allow it to be ready 6 months sooner than Study B.
Benefits to society, if the vaccine works	During the 6 months between now and when the vaccine is ready, it is estimated that: • 1,200,000 people will die from coronavirus. • Millions of people will remain out of work. However, once the vaccine is ready and starts reaching all those who need it in November 2020: • Very few people will die of coronavirus any more. • Daily life and the economy will return to normal.	During the 12 months between now and when the vaccine is ready, it is estimated that: • 2,400,000 people will die from coronavirus. • Millions of people will remain out of work. However, once the vaccine is ready and starts reaching all those who need it in May 2021: • Very few people will die of coronavirus any more. • Daily life and the economy will return to normal.	• With Study A, 1,200,000 lives are saved. • With Study A, people can return to work and daily life can return to normal 6 months sooner.

Figure S1: Example Stimulus, Study 1 (Study A is Challenge Trial in this example)

4

- Challenge Trial N (**80**; 100; 200)
  - N in each condition is calculated as half of this number.
- % in Standard Design that are exposed to coronavirus in their daily lives (**2%**, 5%, 20%). This is calculated in the survey flow and not directly shown to participants.
  - The number of participants who catch coronavirus is calculated as the product of this and the size of the standard trial placebo condition.
- % in Standard Design who die of coronavirus if they are exposed (0.5%, **1%**). This is calculated in the survey flow and not directly shown to participants.
  - The number of participants who die of coronavirus is calculated as the product of this and the number of participants who catch coronavirus, described above.
- How long the Standard Design takes to get a vaccine ready: **12** or 18 months.
  - The date (e.g., **May 2021**), is calculated automatically based on the current date. The number of people in society who die (e.g., **2,400,000**) is calculated by multiplying the number of months until a vaccine is ready by 200,000. 200,000 is a fixed variable for both study designs. We selected 200,000 deaths per month as that is approximately the number of COVID-19 deaths in April 2020, and so therefore represents a likely conservative estimate of likely COVID-19 deaths per month in the months ahead.
- How much faster the Challenge Trial is: 2, 4, or **6** months faster.
  - The date the vaccine is ready if a challenge trial is used is the date above minus this number.
- Which design is described as "Study A" or "Study B". Whichever study was described as "Study A" always was shown in the first column.

We formed these parameters in consultatation with experts in vaccine trial design. The "In the worst-case scenario, they could get an especially bad case of it." language refers to the possibility that "an immune response to a vaccine can predispose an individual to a worse outcome upon infection" [2].

In the "Key Differences" column, the differences are taken based on the randomizations from Studies A and B. We always describe it as possible but unlikely that any participants in the challenge trial would die of coronavirus given findings that the IFR of people ages 18-30 from COVID-19 is 0.03% [6, 7]. (In the largest challenge trial shown to participants (N=200), the probability that zero participants of 100 in the placebo condition would die of coronavirus is $(1 - 0.0003)^{100} = 97\%$.) Table S5, presented later, shows the results are consistent across parameters we used in describing the challenge and standard trials.

After reading this table, we then provided participants with a short summary of the key points. In the example in Figure S1, this would look as follows:

Summary:

- Study **A**:
  - Study participants get coronavirus on their own as they go about their daily lives.
  - [More study participants catch coronavirus and more of them likely die of it. / More study participants are likely to die of coronavirus. / Fewer study participants catch coronavirus, but more would likely die of it. / *If the numbers are equal between the designs, this bullet is omitted.*]
  - It takes longer for the vaccine to be ready, so more people in society generally die of coronavirus.
- Study **B**:
  - Young people volunteer to have scientists expose them to the coronavirus while they are in a medical research center.
  - [Fewer study participants catch coronavirus. It is possible but very unlikely that any study participants would die of it. / More study participants catch coronavirus. It is possible but very unlikely that any study participants would die of it. / *If the numbers are equal between the designs, this bullet is omitted.*]

– The vaccine is ready sooner, so fewer people die of coronavirus in society generally.

Whichever study was randomized to be Study A is always shown first. Which middle bullet in each scenario corresponding to which design has greater or fewer coronavirus cases and deaths in the trial is determined based on the randomization described above. The highlighted sections correspond with the version that would have been shown for the example in Figure S1.

Respondents then answered the questions that constitute our outcome measures. Next, on a separate page, they were asked to answer scenario comprehension questions to ensure they understood the studies. Respondents could not return to the table when answering the scenario comprehension questions. Below we describe the outcome measures and scenario comprehension questions in more detail.

Scenario Comprehension Outcomes

We asked the following scenario comprehension questions to ensure participants, on average, paid attention and understood the survey:

- If the vaccine works, which of the two studies we asked about would result in the vaccine being approved and widely available sooner?
- If the vaccine works, which of the two studies we asked about would result in more people in society generally dying of coronavirus?
- Which of the two studies we asked about involves intentionally exposing participants to coronavirus while they are quarantined in a medical research center?
- Which of the two studies we asked about would result in more people in the study dying of coronavirus?

Responses are coded as "1" if the participant provided the correct answer and "0" if they coded the incorrect answer. Table S2 reports the average rate of correct answers overall and by geography.

There were an unanticipatedly large number of participants in Hong Kong who identified as white, and these participants gave distinctive responses to all the questions, affecting our average characterization of Hong Kong. The 2016 Hong Kong Census estimates that only 0.8% of the Hong Kong population identifies as white [1], so we report the results for self-identified non-white and white Hong Kong participants separately. As Tables S2 and S6 show, self-identified whites in Hong Kong were especially unlikely to understand the scenarios correctly, suggesting this group of self-identified white Hong Kong residents may have been a subset of participants in Hong Kong who were answering the survey carelessly, including the racial identification question itself. Consistent with this interpretation, the median self-identified white participant in Hong Kong spent only 95 seconds reading the main study table (we forced participants to spend at least 60 seconds before they could advance), versus a median of 124 seconds in the rest of the Hong Kong sample and a median of 181 in the sample outside of Hong Kong.

Table S2 also shows the results for the US when weighting the sample to the 2019 US Census population estimates for gender, age, race, and education. We construct these weights using entropy balancing [3].

The p-values in Table S2 are from one-sample t-tests testing the null hypothesis that the rate of correct answers is equal to 0.5, which is what would be expected from random guessing.

The vast majority of participants understood the scenarios. Later, in Table S4, we show that support for the challenge trial is strongest for those who correctly answered all the scenario comprehension questions.

Table S2: Study 1 - Proportions Correctly Answering Scenario Comprehension Questions

	Participants Intentionally Infected In Challenge Correct (0/1)			Vaccine Is Ready Faster With Challenge Correct (0/1)			More Participants Die In Which Study Correct (0/1)			More People in Society Die In Which Study Correct (0/1)			
	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	N
All Participants	0.84	0.82 - 0.85	0	0.83	0.81 - 0.84	0	0.67	0.65 - 0.69	0	0.83	0.81 - 0.84	0	2,988
AUS	0.85	0.80 - 0.89	0	0.83	0.78 - 0.88	0	0.66	0.59 - 0.72	0	0.84	0.79 - 0.89	0	222
CAN	0.85	0.81 - 0.89	0	0.85	0.81 - 0.89	0	0.68	0.63 - 0.73	0	0.83	0.79 - 0.87	0	346
HK (Non-White)	0.75	0.68 - 0.81	0	0.71	0.65 - 0.78	0	0.57	0.50 - 0.64	0.05	0.68	0.61 - 0.74	0	182
HK (White)	0.58	0.37 - 0.78	0.44	0.5	0.29 - 0.71	1	0.54	0.33 - 0.74	0.7	0.58	0.37 - 0.78	0.44	26
NZ	0.82	0.78 - 0.87	0	0.82	0.77 - 0.87	0	0.63	0.57 - 0.69	0	0.83	0.79 - 0.88	0	255
SA	0.88	0.85 - 0.92	0	0.89	0.86 - 0.93	0	0.72	0.67 - 0.78	0	0.91	0.87 - 0.94	0	285
SG	0.79	0.75 - 0.84	0	0.78	0.73 - 0.82	0	0.59	0.54 - 0.65	0	0.79	0.74 - 0.83	0	281
UK	0.84	0.80 - 0.88	0	0.88	0.84 - 0.91	0	0.75	0.70 - 0.80	0	0.87	0.83 - 0.91	0	281
US	0.85	0.83 - 0.87	0	0.83	0.81 - 0.85	0	0.69	0.66 - 0.71	0	0.83	0.81 - 0.86	0	1,110
US, Weighted	0.83	0.81 - 0.85	0	0.82	0.79 - 0.84	0	0.66	0.63 - 0.69	0	0.81	0.78 - 0.83	0	1,110

Primary and Secondary Outcomes

We asked the following outcome measures:

- If you had to choose, which study would you rather have scientists conduct?
- How ethical do you think the studies are? (Asked for both trial designs)
- How scientifically valid do you think the studies are? (Asked for both trial designs)
- If the study found the vaccine worked and it was then approved by the government, how likely would you be to take the vaccine to protect yourself from coronavirus? (Asked for both trial designs)

We pre-specified the first as a primary outcome and the difference between the ratings for the challenge and standard trial designs as secondary outcomes. This difference indicates the extent to which the average respondent judged the challenge trial to be more ethical, scientifically valid, or likelier to be taken than the standard trial, with higher values indicating the challenge trial is more ethical, scientifically valid, or likelier to be taken.

Table S3 shows the overall results and results by geography on each of these dependent variables. We again separately report results for self-identified whites and non-whites in Hong Kong for the reasons described above. The p-value for the primary outcome is from one-sample t-tests testing the null hypothesis that the preference for challenge trials equals the preference for the standard design (0.5), which would indicate an equal number of participants selecting the challenge and standard trial designs. For the other outcomes, the p-values are from one-sample t-tests testing the null hypothesis that the mean is equal to 0, meaning there is no difference between respondents' ratings of the challenge trial and standard design on the outcome measure at hand.

Table S3: Study 1 - Main Results, All Participants and by Participant Geography

	Prefers Scientists Conduct Challenge Study (0/1)			How Ethical (1-4): Challenge Minus Standard			How Scientifically Valid (1-4): Challenge Minus Standard			Would Take Vaccine (1-4): Challenge Minus Standard			
	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	N
All Participants	0.75	0.73 - 0.76	0	0.06	0.02 - 0.10	0.01	0.23	0.20 - 0.27	0	0.14	0.11 - 0.17	0	2,988
AUS	0.79	0.73 - 0.84	0	0.05	-0.10 - 0.19	0.53	0.23	0.12 - 0.34	0	0.07	-0.03 - 0.16	0.15	222
CAN	0.75	0.70 - 0.79	0	-0.05	-0.16 - 0.06	0.39	0.14	0.05 - 0.24	0	0.05	-0.02 - 0.13	0.16	346
HK (Non-White)	0.62	0.54 - 0.69	0	-0.02	-0.20 - 0.17	0.86	0.18	0.03 - 0.32	0.02	0.1	-0.05 - 0.25	0.2	182
HK (White)	0.42	0.22 - 0.63	0.44	0.5	0.08 - 0.92	0.02	-0.27	-0.71 - 0.17	0.22	0	-0.47 - 0.47	1	26
NZ	0.76	0.71 - 0.82	0	0.18	0.04 - 0.31	0.01	0.29	0.18 - 0.40	0	0.25	0.16 - 0.35	0	255
SA	0.83	0.79 - 0.88	0	0.25	0.09 - 0.41	0	0.49	0.37 - 0.61	0	0.37	0.27 - 0.47	0	285
SG	0.66	0.61 - 0.72	0	-0.1	-0.23 - 0.04	0.17	0.17	0.05 - 0.28	0	0.07	-0.04 - 0.17	0.21	281
UK	0.81	0.77 - 0.86	0	0.1	-0.04 - 0.24	0.15	0.23	0.13 - 0.33	0	0.09	-0.00 - 0.18	0.06	281
US	0.74	0.72 - 0.77	0	0.05	-0.02 - 0.11	0.14	0.22	0.16 - 0.28	0	0.15	0.10 - 0.20	0	1,110
US, Weighted	0.73	0.71 - 0.76	0	0.07	0.00 - 0.14	0.04	0.21	0.16 - 0.27	0	0.14	0.09 - 0.19	0	1,110

Table S4 shows that the results are similar across a variety of vulnerable populations, those who answered all the scenario comprehension questions correctly, and various politically relevant subgroups. We also examine a racial subgroup that is defined as non-white in Australia, Canada, New Zealand, the United Kingdom, and the United States, and as non-Chinese in Hong Kong and Singapore. The not college educated subgroup was not pre-specified, but is shown to demonstrate that our conclusions are not driven by the fact that our sample is slightly more educated than the general population.

Table S4: Study 1 - Main Results within Demographic Subgroups

	Prefers Scientists Conduct Challenge Study (0/1)			How Ethical (1-4): Challenge Minus Standard			How Scientifically Valid (1-4): Challenge Minus Standard			Would Take Vaccine (1-4): Challenge Minus Standard			
	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	N
All Participants	0.75	0.73 - 0.76	0	0.06	0.02 - 0.10	0.01	0.23	0.20 - 0.27	0	0.14	0.11 - 0.17	0	2,988
Vulnerable Populations													
Age 65+	0.81	0.77 - 0.85	0	0.17	0.08 - 0.26	0	0.28	0.19 - 0.36	0	0.17	0.10 - 0.25	0	447
Essential Worker	0.69	0.66 - 0.72	0	-0.01	-0.08 - 0.07	0.88	0.19	0.13 - 0.26	0	0.09	0.03 - 0.15	0	883
Non-White (In HK & SG, Non-Chinese)	0.74	0.71 - 0.77	0	0.05	-0.04 - 0.15	0.25	0.23	0.15 - 0.30	0	0.2	0.14 - 0.27	0	748
US: County cases >median	0.74	0.70 - 0.77	0	0.04	-0.05 - 0.14	0.37	0.22	0.14 - 0.30	0	0.17	0.10 - 0.24	0	587
Political Groups (US Only)													
US: Republican	0.7	0.65 - 0.75	0	-0.01	-0.12 - 0.10	0.85	0.18	0.08 - 0.27	0	0.09	0.01 - 0.17	0.02	376
US: Democrat	0.78	0.74 - 0.83	0	0.15	0.03 - 0.27	0.01	0.25	0.15 - 0.36	0	0.21	0.12 - 0.30	0	315
US: Indep/Other	0.75	0.71 - 0.80	0	0.03	-0.09 - 0.14	0.65	0.23	0.14 - 0.33	0	0.16	0.07 - 0.24	0	419
Robustness Checks													
Correctly answered all comprehension questions	0.86	0.85 - 0.88	0	0.17	0.11 - 0.22	0	0.37	0.33 - 0.42	0	0.23	0.19 - 0.26	0	1,529
No college degree	0.77	0.75 - 0.79	0	0.15	0.09 - 0.20	0	0.28	0.23 - 0.33	0	0.18	0.14 - 0.23	0	1,546

Table S5 shows the results are consistent across parameters we used in describing the challenge and standard trials.

Table S5: Study 1 - Main Results, By Randomized Description of Trial Designs

	Prefers Scientists Conduct Challenge Study (0/1)			How Ethical (1-4): Challenge Minus Standard			How Scientifically Valid (1-4): Challenge Minus Standard			Would Take Vaccine (1-4): Challenge Minus Standard			
	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	N
All participants	0.75	0.73 - 0.76	0	0.06	0.02 - 0.10	0.01	0.23	0.20 - 0.27	0	0.14	0.11 - 0.17	0	2,988
Speed of Standard Trial; How Many Months Challenge Accelerates Timeline													
12; Challenge 2 Mo Faster	0.73	0.69 - 0.77	0	0.01	-0.09 - 0.11	0.82	0.23	0.15 - 0.31	0	0.1	0.02 - 0.17	0.01	513
12; Challenge 4 Mo Faster	0.75	0.72 - 0.79	0	0.1	-0.01 - 0.21	0.07	0.27	0.18 - 0.36	0	0.17	0.09 - 0.24	0	477
12; Challenge 6 Mo Faster	0.78	0.74 - 0.81	0	0.07	-0.03 - 0.17	0.19	0.29	0.21 - 0.37	0	0.17	0.10 - 0.25	0	509
18; Challenge 2 Mo Faster	0.72	0.68 - 0.76	0	0.04	-0.06 - 0.14	0.42	0.15	0.06 - 0.24	0	0.13	0.05 - 0.21	0	466
18; Challenge 4 Mo Faster	0.76	0.73 - 0.80	0	0.14	0.04 - 0.24	0.01	0.3	0.22 - 0.38	0	0.13	0.06 - 0.20	0	483
18; Challenge 6 Mo Faster	0.74	0.70 - 0.77	0	-0.01	-0.10 - 0.09	0.88	0.16	0.08 - 0.23	0	0.17	0.10 - 0.23	0	540
% of Placebo Subjects in Standard Trial Who Get Sick, Die													
2% Sick; 0.5% of Sick Die	0.71	0.67 - 0.75	0	-0.16	-0.25 - -0.06	0	0.21	0.13 - 0.29	0	0.06	0.00 - 0.13	0.05	500
2% Sick; 1% of Sick Die	0.75	0.71 - 0.78	0	0.02	-0.08 - 0.12	0.68	0.23	0.15 - 0.31	0	0.17	0.10 - 0.24	0	515
5% Sick; 0.5% of Sick Die	0.75	0.71 - 0.79	0	0.12	0.02 - 0.23	0.02	0.25	0.17 - 0.33	0	0.19	0.12 - 0.26	0	488
5% Sick; 1% of Sick Die	0.76	0.72 - 0.80	0	0.18	0.07 - 0.28	0	0.26	0.18 - 0.35	0	0.16	0.09 - 0.24	0	488
20% Sick; 0.5% of Sick Die	0.73	0.69 - 0.77	0	0.11	0.00 - 0.21	0.04	0.21	0.12 - 0.30	0	0.12	0.04 - 0.20	0.01	476
20% Sick; 1% of Sick Die	0.78	0.74 - 0.82	0	0.08	-0.01 - 0.18	0.09	0.23	0.15 - 0.31	0	0.17	0.09 - 0.24	0	521
Size of Standard Trial													
Standard Trial N = 3000	0.74	0.70 - 0.77	0	0.01	-0.08 - 0.10	0.86	0.27	0.20 - 0.35	0	0.1	0.03 - 0.17	0.01	605
Standard Trial N = 5000	0.77	0.73 - 0.80	0	0.08	-0.02 - 0.17	0.1	0.24	0.17 - 0.31	0	0.16	0.10 - 0.23	0	620
Standard Trial N = 7000	0.75	0.71 - 0.78	0	0.05	-0.04 - 0.14	0.3	0.21	0.14 - 0.29	0	0.16	0.10 - 0.23	0	596
Standard Trial N = 9000	0.75	0.71 - 0.78	0	0.08	-0.01 - 0.18	0.08	0.2	0.12 - 0.28	0	0.14	0.07 - 0.20	0	570
Standard Trial N = 11000	0.73	0.69 - 0.77	0	0.07	-0.02 - 0.17	0.12	0.23	0.16 - 0.31	0	0.16	0.10 - 0.23	0	597

The only case in which participants rate the standard trial as more ethical than the challenge trial is for participants who saw a version of the trial summary where the proportion of subjects in the standard trial who are sick was randomized to the lowest (2%) and the proportion of those who get sick who die was randomized to the lowest (0.5%). Respondents' ratings of the standard trial as more ethical than the challenge trial in this case is driven by participants rating the standard trials as more ethical, not the challenge trial as less ethical, than participants who were randomized to other scenarios. Moreover, it is driven by participants who saw smaller standard trials (Ns of 3,000 or 5,000); in these cases, no participants in the standard trials

die, and fewer participants catch coronavirus than in the challenge trial. It is possible that participants recognized an ethical advantage of the standard trial in this case. However, even these participants still rated the challenge trial similarly ethical as other participants and still favored the challenge trial on average.

Multivariate Regression Results

Figure S2 also visualizes coefficients from a multivariate regression predicting preference for the challenge trial. Variables denoted with (Std.) have been rescaled to standard deviation 1. Note that these coefficients should be interpreted as descriptive, not causal, and that all coefficients are calculated holding constant the other variables shown. The intercept of the regression is 0.78 and the omitted base categories in the regressions are: United States; age 18-24; male; no college degree; employment one of homemaker, student, unable to work, or unemployed since before the coronavirus pandemic; the challenge trial being study A; and average values of the rescaled variables. The coefficients below represent estimated differences from these base categories, holding constant on the other variables in the model.
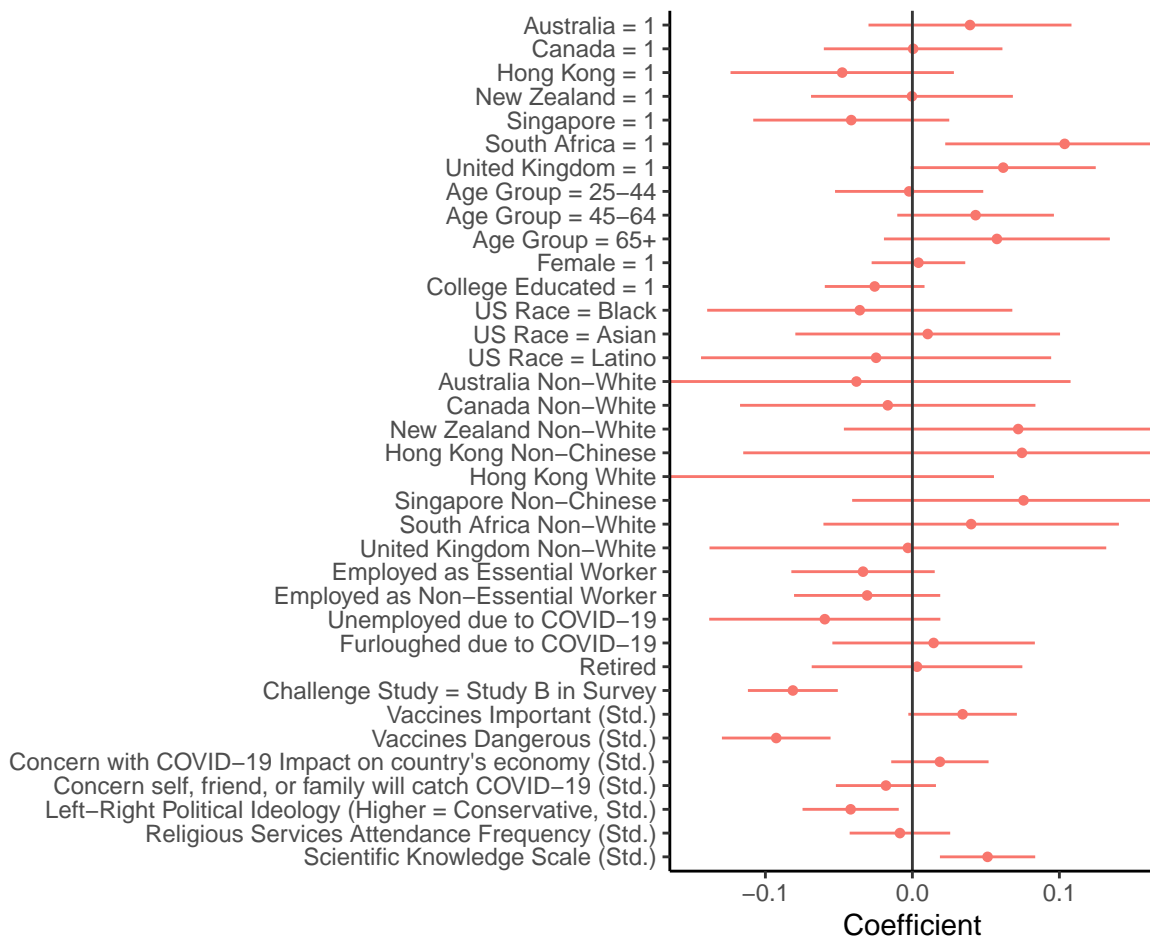

Figure S2: Predictors of Preference for Challenge Trial

Figure S3 shows the results of a multivariate regression regressing the same variables as above on a binary indicator for whether respondents got all the comprehension questions about the challenge study correct. Gender, age, views towards vaccines, and scientific knowledge generally predict comprehension. This regression was not pre-registered.
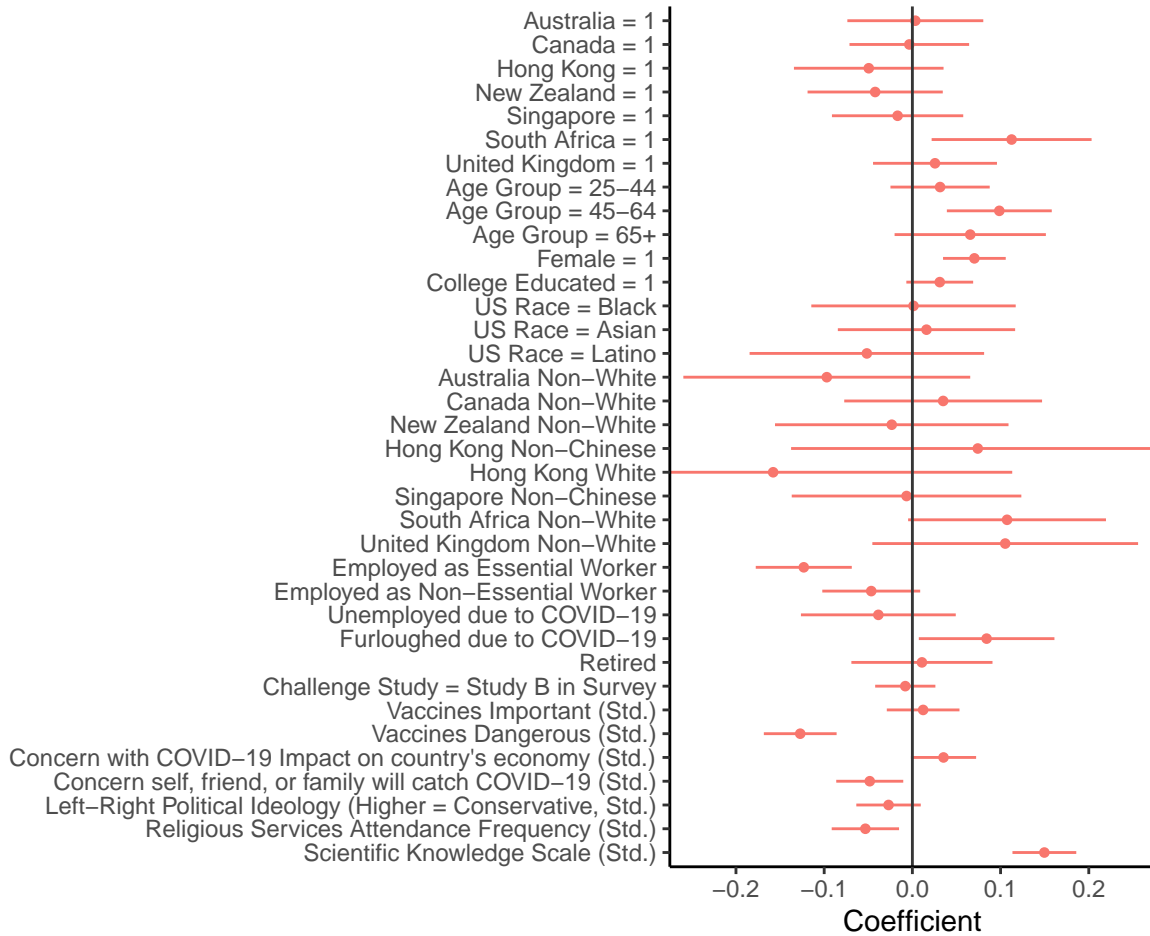
Figure S3: Predictors of Answering Scientific Knowledge Questions Correctly for Challenge Trial

Sensitivity Analysis

We also conduct a sensitivity analysis to determine how unrepresentative our sample would need to be in order to alter our conclusions. In particular, we compute [5]'s γ for the proportion that prefer the challenge study as 2.78 at a p-value of 0.10. This means that our statistical significance for the null hypothesis that challenge trials are equally preferred to standard trials would only no longer reach significance at the 0.10 level were individuals who do not prefer challenge trials to be 2.78 times less likely than individuals who prefer challenge trials to be selected for the survey. For comparison, it is very unusual in social science studies to observe a γ value greater than 2 [4].

Study 2: Integrated Trial

Vaccine Study Design Description

After the preamble quoted earlier, participants were shown a table that looks like the example in Figure S4. All of the highlighted elements were randomized. We did not allow participants to move on from the page describing the trial design until at least 60 seconds had gone by.

Name	Study A	Study B	Key Differences
Study procedures	1. Recruit 9,000 healthy volunteers to participate. 2. Give a few hundred participants the vaccine every week. Stop the study and stop giving participants the vaccine if it is found to be unsafe or unlikely to work. 3. If the vaccine is safe, eventually 4,500 participants will get the vaccine. Another 4,500 will get a placebo.	1. In a preliminary study, recruit 200 healthy people who volunteer to receive the vaccine. After a month, check whether the vaccine has a good chance of working and gather more data on how safe it is. If the vaccine is found to be unsafe or is shown to be unlikely to work, the study stops here. 2. Recruit 9,000 more healthy volunteers to participate. 3. Give 4,500 participants the vaccine. Another 4,500 get a placebo.	• Study B first takes a few months to determine whether the vaccine has a good chance of working and further checks whether it is safe before continuing. Study A immediately starts investigating whether the vaccine actually works, checking to see whether the vaccine is safe and has a good chance of working as it gets started.
Risks to study participants	• While unlikely, if it turns out the vaccine was not safe, approximately 250 people would have received the unsafe vaccine before scientists realized it was unsafe and could have negative side effects.	• If it turns out the vaccine was not safe, 200 people would have received the unsafe vaccine and could have negative side effects.	• While unlikely, if the vaccine is not safe, Study A would expose 50 additional people to it before the problem was found.
Expected time until vaccine ready	If the vaccine works, it would start being distributed widely starting in 10 months, March 2021.	If the vaccine works, it would start being distributed widely in 12 months, May 2021.	• If the vaccine works, Study A would allow it to be ready 2 months sooner.
Benefits to society, if the vaccine works	During the 10 months between now and when the vaccine is ready, it is estimated that: • 2,000,000 people will die from coronavirus. • Millions of people will remain out of work. However, once the vaccine is ready and starts reaching all those who need it in March 2021: • Very few people will die of coronavirus any more. • Daily life and the economy will return to normal.	During the 12 months between now and when the vaccine is ready, it is estimated that: • 2,400,000 people will die from coronavirus. • Millions of people will remain out of work. However, once the vaccine is ready and starts reaching all those who need it in May 2021: • Very few people will die of coronavirus any more. • Daily life and the economy will return to normal.	• With Study A, 400,000 lives are saved. • With Study A, people can return to work and daily life can return to normal 2 months sooner.

Figure S4: Example Stimulus, Study 2 (Study A is Integrated Trial in this example)

The example in Figure S4 shows the results of one particular randomization. All the highlighted numbers in the example were randomized, as detailed below. (These highlights did not appear to survey respondents.) In the below, the bolded numbers correspond with the numbers in the example shown in Figure S4. We

randomized these parameters given uncertainty about how particular vaccine trials might be conducted, to ensure our findings were not sensitive to any of these parameters.

The following elements could have been randomized. The bold corresponds to the example shown in Figure S4.

- Standard Design Trial N (3,000; 5,000; 7,000; **9,000**; 11,000)
  - N in each condition is calculated as half of this number.
- How long the Standard Design takes to get a vaccine ready: **12** or 18 months.
  - The date (e.g., **May 2021**), is calculated automatically based on the current date. The number of people in society who die (e.g., **2,400,000**) is calculated by multiplying the number of months until a vaccine is ready by 200,000. 200,000 is a fixed variable for both study designs. We selected 200,000 deaths per month as that is approximately the number of COVID-19 deaths in April 2020, and so therefore represents a likely conservative estimate of likely COVID-19 deaths per month in the months ahead.
- How much faster the Integrated Design is: **2**, 4, or 6 months faster
  - The date the vaccine is ready if an integrated trial is used is the date above minus this number.
- How many people in the Integrated Design are exposed to the vaccine before it stops, in the case that the vaccine is found to be unsafe (**250**; 400; 1,000)
- Which design is described as "Study A" or "Study B". Whichever study was described as "Study A" always was shown in the first column.

In the "Key Differences" column, the differences are taken based on the randomizations from Studies A and B. Table S9, presented later, shows the results are consistent across parameters we used in describing the integrated and standard trials.

After reading this table, we then provided participants with a short summary of the key points. In the example in Figure S4, this would look as follows:

Summary:

- Study A:
  - Scientists first conduct a study with fewer participants to be extra sure the vaccine is safe and likely to work before continuing.
  - It takes longer for the vaccine to be ready, so more people die of coronavirus in society generally.
- Study B:
  - Scientists give a larger number of participants the vaccine sooner, collecting data as they go to be extra sure the vaccine is safe and has a good chance of working.
  - There is a small chance that the vaccine is unsafe or won't work, in which case 250 more people would have received this ineffective or unsafe vaccine.
  - The vaccine is ready sooner, so fewer people die of coronavirus in society generally.

Whichever study was randomized to be Study A is always shown first.

Respondents then answered the primary outcome measures. Next, on a separate page, they were asked to answer scenario comprehension questions to ensure they understood the studies. Respondents could not return to the table when answering the comprehension questions. Below we describe the outcome measures and comprehension questions in more detail.

Scenario Comprehension Outcomes

We asked the following scenario comprehension questions:

- If the vaccine works, which of the two studies we asked about would result in the vaccine being approved and widely available sooner?
- If the vaccine works, which of the two studies we asked about would result in more people in society generally dying of coronavirus?
- Which of the two studies we asked about involves doing additional safety testing on a smaller group first?

Responses are coded as "1" if the participant provided the correct answer and "0" if they coded the incorrect answer. Table S6 reports the average rate of correct answers overall and by geography. We also show the results for the US when weighting the sample to the 2019 US Census population estimates for gender, age, race, and education, as described above. We again separately report results for self-identified whites and non-whites in Hong Kong for the reasons described above.

The p-value is from one-sample t-tests testing the null hypothesis that the rate of correct answers is equal to 0.5, which is what would be expected from random guessing.

The vast majority of participants understood the scenarios. Later, in Table S8, we show that support for the integrated trial is strongest for those who correctly answered all the scenario comprehension.

Table S6: Study 2 - Proportions Correctly Answering Scenario Comprehension Questions

	Vaccine Is Ready Faster With Integrated Correct (0/1)			More People in Society Die In Which Study Correct (0/1)			Standard Involves Additional Testing Correct (0/1)			
	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	N
All Participants	0.85	0.83 - 0.86	0	0.81	0.79 - 0.82	0	0.75	0.73 - 0.76	0	2,932
AUS	0.84	0.79 - 0.88	0	0.81	0.76 - 0.85	0	0.76	0.71 - 0.81	0	278
CAN	0.86	0.82 - 0.90	0	0.79	0.75 - 0.84	0	0.75	0.71 - 0.80	0	341
HK (Non-White)	0.76	0.70 - 0.82	0	0.73	0.67 - 0.79	0	0.64	0.58 - 0.71	0	194
HK (White)	0.55	0.31 - 0.79	0.67	0.65	0.42 - 0.88	0.19	0.35	0.12 - 0.58	0.19	20
NZ	0.86	0.81 - 0.90	0	0.82	0.77 - 0.87	0	0.78	0.73 - 0.83	0	243
SA	0.93	0.90 - 0.96	0	0.87	0.83 - 0.91	0	0.78	0.73 - 0.83	0	263
SG	0.79	0.74 - 0.84	0	0.8	0.75 - 0.85	0	0.69	0.63 - 0.75	0	239
UK	0.88	0.84 - 0.92	0	0.83	0.79 - 0.88	0	0.79	0.74 - 0.84	0	284
US	0.85	0.83 - 0.87	0	0.79	0.77 - 0.82	0	0.76	0.74 - 0.79	0	1,070
US, Weighted	0.85	0.83 - 0.87	0	0.8	0.78 - 0.83	0	0.76	0.73 - 0.78	0	1,070

Primary and Secondary Outcomes

We asked the following outcome measures:

- If you had to choose, which study would you rather have scientists conduct?
- How ethical do you think the studies are?
- How scientifically valid do you think the studies are?
- If the study found the vaccine worked and it was then approved by the government, how likely would you be to take the vaccine to protect yourself from coronavirus?

We pre-specified the first as a primary outcome and the difference between the ratings for the integrated and standard trial designs as secondary outcomes.

Table S7 shows the overall results and results by geography on each of these dependent variables. We again separately report results for self-identified whites and non-whites in Hong Kong for the reasons described above. The p-value for the primary outcome is from one-sample t-tests testing the null hypothesis that an equal number of participants selecting the integrated and standard trial designs (0.5). For the other outcomes, the p-values are from one-sample t-tests testing the null hypothesis that the mean is equal to 0, meaning

there is no difference respondents' ratings of the integrated trial and standard design on the outcome measure at hand.

Even though participants on average are more likely to select the integrated trial, the averages of the secondary outcomes are slightly higher for the standard design because those who select the standard design gave more extreme responses. For example, participants who selected the integrated trial rated the integrated trial as equally ethical (only 0.01 scale points less ethical on average), whereas those who favored the standard trial rated the integrated trial as 0.62 scale points less ethical on average. However, 58% of respondents who did not prefer the integrated trial still said they thought the integrated trial was "probably" or"definitely ethical."

Table S7: Study 2 - Main Results, All Participants and by Participant Geography

	Prefers Scientists Conduct Integrated Study (0/1)			How Ethical (1-4): Integrated Minus Standard			How Scientifically Valid (1-4): Integrated Minus Standard			Would Take Vaccine (1-4): Integrated Minus Standard			
	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	N
All Participants	0.63	0.61 - 0.65	0	-0.24	-0.27 - -0.21	0	-0.05	-0.07 - -0.03	0	-0.05	-0.08 - -0.03	0	2,932
AUS	0.64	0.58 - 0.69	0	-0.27	-0.35 - -0.19	0	-0.05	-0.12 - 0.02	0.18	-0.06	-0.14 - 0.03	0.19	278
CAN	0.66	0.61 - 0.71	0	-0.24	-0.32 - -0.16	0	-0.05	-0.11 - 0.02	0.15	-0.04	-0.10 - 0.03	0.3	341
HK (Non-White)	0.52	0.44 - 0.59	0.67	-0.29	-0.43 - -0.15	0	-0.05	-0.16 - 0.06	0.36	-0.11	-0.24 - 0.02	0.1	194
HK (White)	0.25	0.04 - 0.46	0.02	-0.4	-0.72 - -0.08	0.02	0.05	-0.49 - 0.59	0.85	-0.4	-0.87 - 0.07	0.09	20
NZ	0.6	0.53 - 0.66	0	-0.15	-0.25 - -0.05	0	-0.03	-0.11 - 0.05	0.47	-0.04	-0.12 - 0.05	0.38	243
SA	0.66	0.60 - 0.72	0	-0.26	-0.37 - -0.14	0	-0.06	-0.15 - 0.03	0.17	-0.02	-0.12 - 0.07	0.64	263
SG	0.5	0.44 - 0.57	0.95	-0.37	-0.49 - -0.25	0	-0.05	-0.15 - 0.04	0.28	-0.11	-0.22 - 0.00	0.06	239
UK	0.7	0.65 - 0.75	0	-0.27	-0.36 - -0.18	0	-0.04	-0.10 - 0.03	0.24	-0.04	-0.11 - 0.04	0.36	284
US	0.65	0.62 - 0.68	0	-0.18	-0.23 - -0.14	0	-0.05	-0.09 - -0.01	0.01	-0.02	-0.06 - 0.02	0.38	1,070
US, Weighted	0.65	0.62 - 0.68	0	-0.19	-0.23 - -0.14	0	-0.06	-0.09 - -0.02	0.01	-0.05	-0.09 - -0.00	0.03	1,070

Table S8 shows that the results are similar across a variety of vulnerable populations, those who answered all the scenario comprehension questions correctly, and various politically relevant subgroups. As described above, we also examine a racial subgroup of non-white in Australia, Canada, New Zealand, the United Kingdom, and the United States, and as both non-Chinese and non-white in Hong Kong and Singapore. The not college educated subgroup was not pre-specified, but is shown to demonstrate that our conclusions are not driven by the fact that our sample is slightly more educated than the general population.

Table S8: Study 2 - Main Results By Demographic Groups

	Prefers Scientists Conduct Integrated Study (0/1)			How Ethical (1-4): Integrated Minus Standard			How Scientifically Valid (1-4): Integrated Minus Standard			Would Take Vaccine (1-4): Integrated Minus Standard			
	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	N
All Participants	0.63	0.61 - 0.65	0	-0.24	-0.27 - -0.21	0	-0.05	-0.07 - -0.03	0	-0.05	-0.08 - -0.03	0	2,932
Vulnerable Populations													
Age 65+	0.74	0.70 - 0.78	0	-0.12	-0.18 - -0.06	0	-0.01	-0.06 - 0.04	0.67	0.04	-0.02 - 0.10	0.17	507
Essential Worker	0.57	0.54 - 0.60	0	-0.23	-0.28 - -0.17	0	-0.07	-0.12 - -0.02	0	-0.04	-0.10 - 0.01	0.1	880
Non-White (In HK & SG, Non-Chinese)	0.55	0.52 - 0.59	0	-0.29	-0.35 - -0.22	0	-0.09	-0.15 - -0.03	0	-0.1	-0.17 - -0.04	0	728
US: County cases >median	0.64	0.60 - 0.68	0	-0.19	-0.26 - -0.13	0	-0.06	-0.12 - -0.00	0.04	-0.04	-0.10 - 0.02	0.21	550
Political Groups (US Only)													
US: Republican	0.68	0.63 - 0.73	0	-0.14	-0.21 - -0.06	0	-0.03	-0.09 - 0.04	0.43	-0.05	-0.12 - 0.02	0.19	358
US: Democrat	0.63	0.58 - 0.68	0	-0.15	-0.23 - -0.07	0	-0.04	-0.11 - 0.03	0.26	-0.01	-0.07 - 0.06	0.85	331
US: Indep/Other	0.65	0.60 - 0.69	0	-0.27	-0.35 - -0.19	0	-0.1	-0.17 - -0.03	0.01	-0.08	-0.15 - -0.00	0.04	381
Robustness Checks													
Correctly answered all comprehension questions	0.68	0.66 - 0.70	0	-0.29	-0.33 - -0.26	0	-0.08	-0.10 - -0.05	0	-0.05	-0.08 - -0.02	0	1,745
No college degree	0.64	0.61 - 0.66	0	-0.2	-0.24 - -0.15	0	-0.03	-0.06 - 0.00	0.08	-0.04	-0.07 - 0.00	0.05	1,496

Table S5 shows the results are consistent across parameters we used in describing the integrated and standard trials. Unsurprisingly, respondents were slightly less likely to favor the integrated design when greater numbers of study participants receive the vaccine before a safety or immunogenicity problem with the vaccine could be detected.

Table S9: Study 2 - Main Results By Randomized Description of Trial Designs

	Prefers Scientists Conduct Integrated Study (0/1)			How Ethical (1-4): Integrated Minus Standard			How Scientifically Valid (1-4): Integrated Minus Standard			Would Take Vaccine (1-4): Integrated Minus Standard			
	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	Mean	95% CI	p	N
All Participants	0.63	0.61 - 0.65	0	-0.24	-0.27 - -0.21	0	-0.05	-0.07 - -0.03	0	-0.05	-0.08 - -0.03	0	2,932
Speed of Standard Trial; How Many Months Integrated Accelerates Timeline													
12; Integrated 2 Mo Faster	0.57	0.53 - 0.61	0	-0.23	-0.30 - -0.16	0	-0.04	-0.09 - 0.01	0.15	-0.13	-0.19 - -0.06	0	501
12; Integrated 4 Mo Faster	0.63	0.59 - 0.68	0	-0.25	-0.32 - -0.18	0	-0.03	-0.08 - 0.03	0.36	-0.07	-0.13 - -0.00	0.04	495
12; Integrated 6 Mo Faster	0.69	0.65 - 0.73	0	-0.21	-0.29 - -0.14	0	-0.04	-0.10 - 0.02	0.16	-0.02	-0.08 - 0.04	0.55	509
18; Integrated 2 Mo Faster	0.58	0.53 - 0.62	0	-0.27	-0.34 - -0.19	0	-0.05	-0.11 - 0.02	0.14	-0.04	-0.11 - 0.04	0.33	451
18; Integrated 4 Mo Faster	0.64	0.60 - 0.68	0	-0.25	-0.32 - -0.18	0	-0.03	-0.09 - 0.03	0.33	-0.04	-0.10 - 0.02	0.19	490
18; Integrated 6 Mo Faster	0.65	0.60 - 0.69	0	-0.21	-0.28 - -0.14	0	-0.11	-0.17 - -0.05	0	-0.03	-0.09 - 0.03	0.33	486
# Of Integrated Subjects Who Get Vaccine Before Trial Stopped, if Vaccine Unsafe or Not Immunogenic													
250	0.67	0.64 - 0.70	0	-0.2	-0.25 - -0.15	0	-0.02	-0.06 - 0.02	0.35	-0.02	-0.07 - 0.02	0.31	992
400	0.61	0.58 - 0.64	0	-0.24	-0.29 - -0.19	0	-0.04	-0.08 - 0.01	0.09	-0.05	-0.09 - -0.00	0.05	943
1,000	0.6	0.57 - 0.63	0	-0.27	-0.32 - -0.21	0	-0.09	-0.13 - -0.05	0	-0.09	-0.14 - -0.04	0	997

Multivariate Regression Results

Figure S5 also visualizes coefficients from a multivariate regression predicting preference for the integrated trial. Variables denoted with (Std.) have been rescaled to standard deviation 1. Note that these coefficients should be interpreted as descriptive, not causal, and that all coefficients are calculated holding constant the other variables shown. The intercept of the regression is 0.64 and the omitted base categories in the regressions are: United States; age 18-24; male; no college degree; employment one of homemaker, student, unable to work, or unemployed since before the coronavirus pandemic; the integrated trial being study A; and average values of the rescaled variables. The coefficients below represent estimated differences from these base categories, holding constant the other variables in the model.

Figure S6 shows the results of a multivariate regression regressing the same variables as above on a binary indicator for whether respondents got all the comprehension questions about the integrated study correct. Gender, age, views towards vaccines, and scientific knowledge generally predict comprehension. This regression was not pre-registered.

Sensitivity Analysis

We also conduct a sensitivity analysis to determine how unrepresentative our sample would need to be in order to alter our conclusions. In particular, we compute [5]'s γ for the proportion that prefer the integrated study as 1.61 at a p-value of 0.10. This means that our statistical significance for the null hypothesis that integrated trials are equally preferred to standard trials would only no longer reach significance at the 0.10 level were individuals who do not prefer integrated trials to be 1.61 times less likely than individuals who prefer integrated trials to be selected for the survey. For comparison, it is very unusual in social science studies to observe a γ value greater than 2 [4].
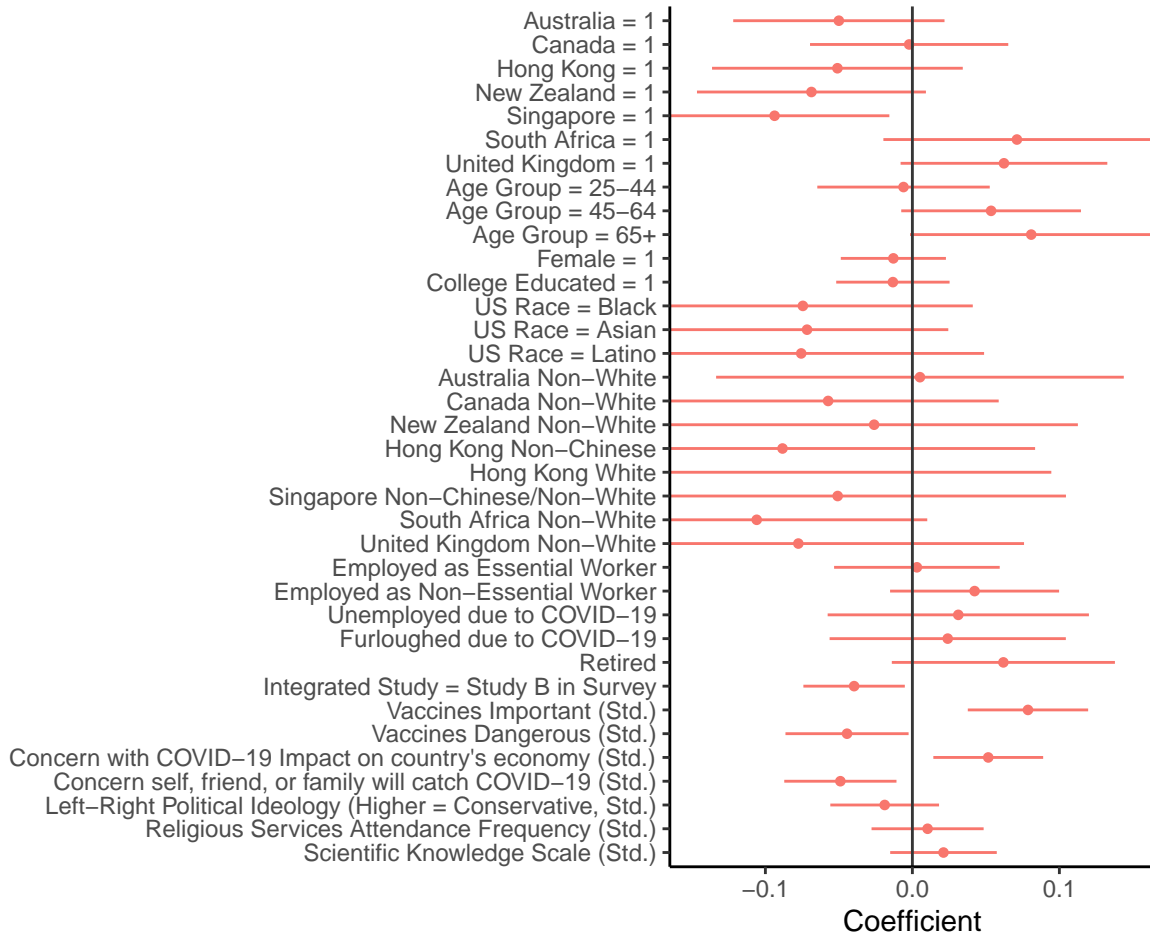
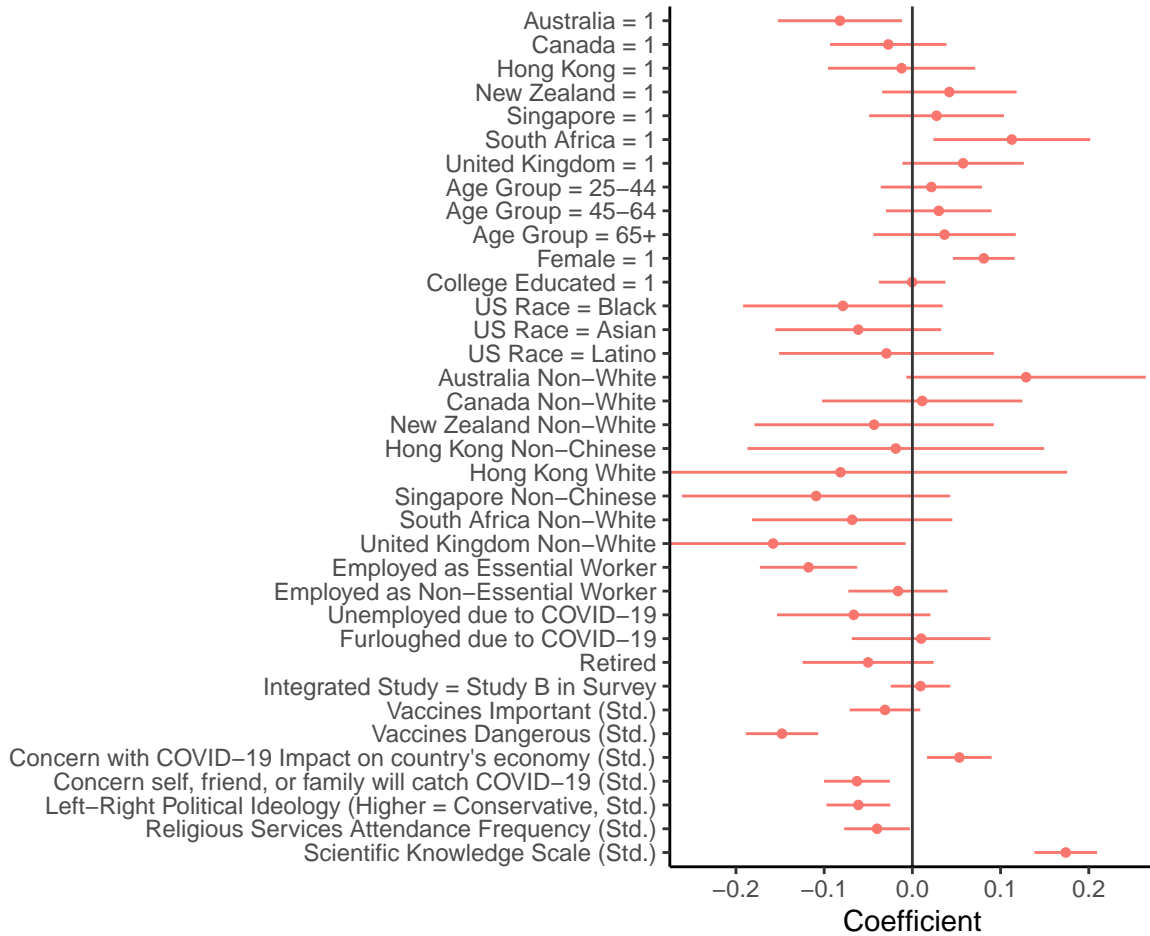Figure S5: Predictors of Preference for Integrated Trial

Figure S6: Predictors of Answering Scientific Knowledge Questions Correctly for Integrated Trial

Pre-Registration

Departures from Pre-Registration

We made only minor departures from our pre-registered analysis in our pre-analysis plan:

- Given that participants to the survey were more likely to be college educated than the general population, we present results for "non-college educated" participants as a subgroup to show our results are robust among this group, which we had not pre-specified.
- There were an unanticipatedly large number of participants in Hong Kong who identified as white, and these participants gave distinctive responses to all the questions, affecting our average characterization of Hong Kong. The 2016 Hong Kong Census estimates that only 0.8% of the Hong Kong population identifies as white [1], so we report the results of self-identified non-white and white Hong Kong participants separately. As Tables S2 and S6 show, self-identified whites in Hong Kong were especially unlikely to understand the scenarios correctly, suggesting this group of self-identified white Hong Kong residents may have been a subset of participants in Hong Kong who were answering the survey carelessly, including the racial identification question itself. Consistent with this interpretation, the median self-identified white participant in Hong Kong spent only 95 seconds reading the main study table (we did not allow participants to move on from the page describing the trial design until at least 60 seconds had gone by), versus a median of 124 seconds in the rest of the Hong Kong sample and a median of 181 in the sample outside of Hong Kong.

Pre-Registration Document

Our pre-registration document appears on the following pages. It was filed prior to the collection of our survey data.

Pre-Registration and Pre-Analysis Plan:
Public Perception of Ethical Trade-offs in COVID-19 Vaccine Trial Design

We will be conducting an online survey to gauge how the public weighs various ethical trade-offs related to the design of vaccine trials. We will be conducting this survey in the United States, United Kingdom, Australia, Canada, New Zealand, Singapore, Hong Kong, and South Africa using the online survey provider Lucid. We aim for a sample size of 500 respondents per country and 2,000 in the United States. Respondents must take the survey on a desktop and be able to read English.

In the survey, we first ask two attention check questions. If respondents fail those attention checks, they are removed from the survey.

We then randomly assign respondents to one of two studies:
- Challenge Trial vs. Standard Design
- Integrated Phase 2/3 vs. Standard Design

In the Challenge Trial vs. Standard Design study, we randomize the following features:
- Standard Design Trial N (3,000; 5,000; 7,000; 9,000; 11,000)
- Challenge Trial N (80; 100; 200)
- % in Standard Design that are exposed to coronavirus in their daily lives (2%, 5%, 20%)
- % in Standard Design who die of coronavirus if they are exposed (0.5%, 1%)
- How long the Standard Design takes to get a vaccine ready: 12 or 18 months
- How much faster the Challenge Trial is: 2, 4, or 6 months faster
- Which design is described as "Study A" or "Study B"

In the Integrated Phase 2/3 vs. Standard Design study, we randomize the following features:
- Standard Design Trial N (3,000; 5,000; 7,000; 9,000; 11,000)
- How long the Standard Design takes to get a vaccine ready: 12 or 18 months
- How much faster the Integrated Design is: 2, 4, or 6 months faster
- How many people in the Integrated Design are exposed to the vaccine before it stops, in the case that the vaccine is found to be unsafe (250; 400; 1,000)
- Which design is described as "Study A" or "Study B"

For each study, we will ask the following outcome measures:
- Primary outcome
  - If you had to choose, which study would you rather have scientists conduct? Study A; Study B
    - This will be coded as 1 for Challenge/Integrated and 0 for Standard
- Secondary outcomes
  - How ethical do you think the studies are? Asked for both designs
    - Definitely ethical (4); Probably ethical (3); Probably unethical (2); Definitely unethical (1)

- - We will analyze this outcome by taking the difference between Challenge/Integrated minus Standard
  - We will also report the frequencies for the individual variables
  - How scientifically valid do you think the studies are? Asked for both Study A and Study B
    - Very valid (4); Somewhat valid (3); Somewhat invalid (2); Very invalid (1)
    - We will analyze this outcome by taking the difference between Challenge/Integrated minus Standard
    - We will also report the frequencies for the individual variables
  - If the study found the vaccine worked and it was then approved by the government, how likely would you be to take the vaccine to protect yourself from coronavirus? Asked for both Study A and Study B
    - Very likely (4); Somewhat likely (3); Somewhat unlikely (2); Very unlikely (1)
    - We will analyze this outcome by taking the difference between Challenge/Integrated minus Standard
    - We will also report the frequencies for the individual variables

For each study, we will also ask the following factual understanding questions to ensure respondents, on average, paid attention and understood the survey:
- If the vaccine works, which of the two studies we asked about would result in the vaccine being approved and widely available sooner?
- If the vaccine works, which of the two studies we asked about would result in more people in society generally dying of coronavirus?
- These questions will only be asked for the Challenge Trial vs. Standard Design study:
  - Which of the two studies we asked about involves intentionally exposing participants to coronavirus while they are quarantined in a medical research center?
  - Which of the two studies we asked about would result in more people in the study dying of coronavirus?
    - Note: This question has a "Neither" option because in some randomizations, the number is the same.
- This question will only be asked for the Integrated vs. Standard Design study:
  - Which of the two studies we asked about involves doing additional safety testing on a smaller group first?
- Each factual understanding variable will be recoded to have 1 for the correct answer and 0 for the incorrect answer based on the randomization.

We will conduct the below analyses for each study (challenge and integrated). For each average we describe below, we will perform a one-sample t-test, testing the null hypothesis that the challenge/integrated and standard designs are equal; this implies a null of 0.5 for the "If you had to choose" variable and a null of 0 for the secondary outcomes.
- Average value for each outcome, overall and by country
- Subgroups of primary interest are listed below. Our goal for subgroup analyses is to demonstrate the consistency of the findings across a) randomized descriptions of trial designs and b) salient

social cleavages, especially among vulnerable populations and politically relevant groups. With this in mind, we will compute the average value and perform the t-tests mentioned above among participants in each of the subgroups mentioned below. We will only examine subgroups that are at least N=50 in size.

- ○ Demographic groups
    - ■ Only people 65 and over, given they are at highest risk for serious complications or death from coronavirus
    - ■ Only participants who understood all the factual understanding questions correctly
    - ■ Only participants who say they are "essential workers"
    - ■ Racial minorities. We will measure this as follows:
        - ● US / UK / Australia / NZ / South Africa / Canada: those who do not select "White" to the race/ethnicity question
        - ● Singapore and HK: Those who do not select "Chinese" to a race/ethnicity question
    - ■ Generally speaking, do you consider yourself a...Democrat/Republican/Independent/Other Party. We will create indicators for each, pooling Independents and Other Party into one category. (This analysis will be done for US respondents only.)
    - ■ In a US county with cumulative COVID cases per capita above the median. To calculate COVID cases per capita, we will compute county population using 2019 Census population estimates (https://www2.census.gov/programs-surveys/popest/tables/2010-2019/counties/totals/co-est2019-annres.xlsx) and COVID cases determined on the date of the launch of the survey, using the *New York Times* data at https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv. (This analysis will be done for US respondents only.)
- ○ Randomized descriptions of trial designs
    - ■ By the number of months the Standard design takes and how much faster the Challenge/Integrated designs will be, as a 2x3 table with 6 separate statistics (reported separately for Integrated and Challenge)
    - ■ Average value for each outcome by the death rate and sick rate in the trial, as a 2x3 table with 6 separate statistics (Challenge only)
    - ■ Average value for each outcome by number of people who get the vaccine before it is determined to be unsafe (Integrated only)
- ● For each outcome, we will also report a regression to estimate which demographics predict support. We may also report raw means of outcomes within demographic categories. We will include the following predictors, all as linear predictors unless specified otherwise:
    - ○ In your opinion, how important is it that parents get their children vaccinated? Extremely important (5); Very important (4); Somewhat important (3); Not very important (2); Not at all important (1)
    - ○ Do you think vaccines are more dangerous than the diseases they are designed to prevent, or not? Yes (3); Unsure (2); No (1)

- How concerned are you about the effect of the coronavirus on the country's economy? Very concerned (4); Somewhat concerned (3); Not very concerned (2); Not at all concerned (1)
- How concerned are you that you, someone in your family, or someone else you know will become infected with coronavirus? Very concerned (4); Somewhat concerned (3); Not very concerned (2); Not at all concerned (1)
- What is your year of birth? Recoded as age and groups into bins: 18-24; 25-44; 45-64; 65+, each analyzed as an indicator variable
- Which of the following best describes your gender? 1 = female; 0 = all other
- What is the highest level of education you have completed? 1 = college educated or above; 0 = all other
- In political matters, people talk of "the left" or "liberal" and "the right" or "conservative". How would you place your views on this scale, generally speaking? Coded from 1 (Liberal) to 10 (Conservative)
- What is your current employment status, and are you considered an "essential worker" during this pandemic? Indicators for employed as an essential worker; employed as a non-essential worker; unemployed due to COVID; furloughed due to COVID; and retired.
- Apart from weddings and funerals, about how often did you usually attend religious services last year? More than once a week (5); Once a week (4); Once a month (3); Only on special holy days (2); Once (1); Never (0)
- Scientific knowledge. We ask respondents if they know "Which kind of waves is used to make and receive cellphone calls?" and "Ocean tides are created by which of the following?". Respondents get a 2 if they answer both correctly; a 1 if they answer one correctly; and a 0 if they answer none correctly
- An indicator for every country.
- Race/ethnicity: We will create indicators for US Black, US Asian, US Latino, and for non-white in each of the UK, Australia, NZ, South Africa, and Canada, and for non-Chinese in Singapore and Hong Kong.
- An indicator for whether the integrated/challenge study was randomized to be "Study A" or "Study B".

For a separate research question, we will also analyze the mean for the question examining preferences about a post-challenge trial safety study involving either 3,000 or 1 million people, both overall and among those 65+ only. We may report both these results separately.

We may also conduct qualitative analyses of the open-ended responses to the questions asking respondents why they gave the answers about the ethics they did. Details of how we will conduct this analysis are not pre-registered.

Our primary analyses will be unweighted. As a robustness check, we will also present results for the United States using weights. For this analysis, we will weight to the ACS on age, gender, education, and

race using the ebalance package in Stata. We will compute the weighted mean using the wtd.t.test function from the weights package in R. Our analysis will assume the weights are fixed.

References

[1] Census and Statistics Department, Government of the Hong Kong Special Administrative Region of the People's Republic of China. Main tables, table a104. Available at https://www.bycensus2016.gov.hk/en/bc-mt.html.

[2] Douglas R Green. Sars-cov2 vaccines: Slow is fast. *Science Advances*, 2020.

[3] Jens Hainmueller. Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, 20(1):25–46, 2012.

[4] Luke Keele. An overview of rbounds: An r package for rosenbaum bounds sensitivity analysis with matched data. *White Paper*, 2010.

[5] Paul R Rosenbaum. Observational studies. 2020.

[6] Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick GT Walker, Han Fu, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 2020.

[7] World Health Organization. Key criteria for the ethical acceptability of covid-19 human challenge studies. Technical report, World Health Organization, 2020.