medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

PREPRINT ARTICLE

Exploring Risks of Human Challenge Trials for COVID-19

David Manheim, PhD[*,1,2] | Witold Więcek, PhD[*,1,3] | Virginia Schmit, PhD[1] | Josh Morrison[1] | 1Day Sooner Research Team

[*]Equally Contributing Authors

[1]1DaySooner, Delaware, United States

[2]Health and Risk Communication Research Center, University of Haifa, School of Public Health, Haifa, Israel

[3]WAW Statistical Consulting Ltd, United Kingdom

Correspondence
David Manheim, PhD
Email: media@1daysooner.org

Human Challenge Trials (HCTs) are a potential method to accelerate development of vaccines and therapeutics. However, HCTs for COVID-19 pose ethical and practical challenges, in part due to the unclear and developing risks. In this paper, we introduce an interactive model for exploring some risks of a SARS-COV-2 dosing study, a prerequisite for any COVID-19 challenge trials. The risk estimates we use are based on a Bayesian evidence synthesis model which can incorporate new data on infection fatality rates (IFRs) to patients, and infer rates of hospitalization. We have also created a web tool to explore risk under different study design parameters and participant scenarios. Finally, we use our model to estimate individual risk, as well as the overall mortality and hospitalization risk in a dosing study.

Based on the Bayesian model we expect IFR for someone between 20 and 30 years of age to be 17.5 in 100,000, with 95% uncertainty interval from 12.8 to 23.6. Using this estimate, we find that a simple 50-person dosing trial using younger individuals has a 99.1% (95% CI: 98.8% to 99.4%) probability of no fatalities, and a 92.8% (95% CI: 90.3% to 94.6%) probability of no cases requiring hospitalization. However, this IFR will be reduced in an HCT via screening for comorbidities, as well as providing medical care and aggressive treatment for any cases which occur, so that with stronger assumptions, we project the risk to be as low as 3.1 per 100,000, with a 99.85% (95% CI: 99.7% to 99.9%) chance of no fatalities, and a 98.7% (95% CI: 97.4% to 99.3%) probability of no cases requiring hospitalization.

KEYWORDS
Human Challenge Trial, COVID-19, Informed Consent, Bayesian, Meta-analysis, Risk Communication, Interactive Models

Abbreviations: HCT, Human Challenge Trial; IFR, infection fatality rate

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

1 | INTRODUCTION

As of November 15, 2020, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has led to over 54 million confirmed infections worldwide and over 1.3 million deaths. Currently, there are no clinically-approved vaccines against COVID-19. McKinsey estimates that global vaccine production will reach one billion doses by the end of 2020, and eight billion by 2021 assuming the most advanced vaccine candidates pass clinical trials on time[2]–a thin margin of error.

Timely development of new pharmacological therapeutics and vaccines will be necessary to manage disease burden and impact of COVID-19. Different paths to vaccines can exist, and more traditional clinical trials are slow and require large subject populations to discern therapeutic effects[18]. In order to accelerate testing, several authors and institutions have proposed intentionally exposing human subjects to SARS-CoV-2 to test novel interventions; to this date, nearly forty thousands of people have expressed interest in volunteering for this task[1]. Such human challenge trials (HCTs) have been useful in the past[19] to develop vaccines and treatments for other infectious diseases such as Malaria, cholera, respiratory syncytial virus (RSV) [35, 13], influenza [41], and Dengue fever [20]. However, two major problems currently stand in the way of conducting HCTs for SARS-CoV-2 infection. First, given lack of rescue therapies and our limited understanding of COVID-19's risks, it is difficult to weigh the likely impacts of these studies on volunteers against the benefit to society, or to obtain informed consent from volunteers[33]. Second, we do not know what viral dose of SARS-CoV-2 should be given to volunteers.

To help address both concerns, we developed a model to help assess the risks participants will face in a hypothetical dosing study for COVID-19. This model uses data from a non-systematic review of data on COVID-19 risks (mortality and infection rates) and describes risks for individuals as well as the overall study risk. As both clarification on viral dose and viral risk are essential before starting HCTs, this work can help inform policymakers and potential volunteers about some risks concerning the process of using HCTs to accelerate vaccine and therapeutic development.

2 | METHODOLOGY

We developed a three-component tool to understand and explain the relevant risks. The first component quantifies risk of COVID-19 mortality by using a Bayesian evidence synthesis model; the second uses that estimate, along with other data on gender-specific mortality and hospitalization risks to simulate the risk of a study with given characteristics; the third is a front end tool for allowing interactive exploration of risks from a study or to an individual.

2.1 | Bayesian evidence synthesis model

We use a Bayesian meta-analysis approach to obtain an estimate of mortality risk. This form of modelling combines different sources of evidence with varying statistical power to obtain posterior distribution for mortality risk in different populations. In our case, we use age-specific, location-specific death and prevalence data to generate an estimate of the infection fatality rate (IFR), and estimates of relative risk in healthy individuals (vs the general population) to understand risk reduction in individuals who would participate in an HCT.

We use Bayesian methods because they allow us to best account for heterogeneity in IFRs across different populations (age groups, different countries and regions). Characterising this heterogeneity is important when assessing possible reductions in IFRs. For example, it can be argued that an HCT can use screening and provide medical care to

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

achieve a rate of IFRs which is at least as low[1] as the region or country with the lowest IFRs in our data.

Although existing statistical packages for meta-analysis (both Bayesian and frequentist) could easily be used to model event rates such as IFR[44, 8], these models may not work when no deaths are observed, as is often the case for COVID-19 in younger populations. To address this, we use death data and estimates of prevalence as inputs instead of IFRs. We then construct a new, reproducible model for IFRs.

Methodological details of the model are described in the Appendix. [2] We assume that the fixed effect of age and random effects of location on IFR are on logit scale, as is typical for meta-analysis models of binary data.

However, while our IFR estimates capture average risks within different age groups and even heterogeneity across regions or countries, they still refer to the general population (of a given age, in a given location). A prospective HCT participant would be screened for health issues and comorbidities, further reducing the risk in comparison to the members of general population. To account for this, we perform an additional analysis using OpenSAFELY[43], a large observational dataset on COVID-19 mortality factors that includes comorbidity and age information, as well as data on gender, comparing the risk in general population to lower-risk subpopulation[3]. Similarly to adjustment for heterogeneity across general populations, the adjustment for screening can be turned on or off in our tool.

Input data for the Bayesian IFR model is based on a non-systematic review of the literature and earlier meta-analyses, particularly by Levin *et al* [21], but we have also opportunistically included other studies, and data from other official sources as detailed in the appendix. However, the list of studies is not fixed, since new and better characterized datasets are becoming available over time. For this reason, we have made sure that incorporating additional prevalence and death data from newer studies and/or updated data sets is straightforward. We will continually update the model to assure that any estimates provided to participants or used for decisions include all relevant data, rather than only using data that was available when the analysis was first performed.

We use the age- and gender-specific data from Salje et al. [36] for the rate of death of hospitalized patients to impute the relative risk of hospitalization based on our meta-analysis for mortality risk.

Our model is available publicly, together with input data and source code for the tool and under an open license, at `https://github.com/1DaySooner/RiskModel`.

[1] This is an assumption, and we allow users of the tool to choose to use the overall estimate across studies, or the expected lower risk.

[2] Briefly, let k index populations and loc_k their locations (countries, regions). Let d_k be observed deaths for data point k; $\mu_k^{(p)}$ the reported mean prevalence and $\sigma_k^{(p)}$ its standard error, on logit scale; n_k the total population; X is a vector of median ages, expressed in decades and centered at 25 years. The hierarchical model is as follows:

$$d_k \sim \text{Binomial}(n_k, p_k IFR_k) \tag{1}$$

$$\text{logit}(p_k) \sim \mathcal{N}(\mu_k^{(p)}, (\sigma_k^{(p)})^2) \tag{2}$$

$$\text{logit}(IFR_k) = \theta_{\text{loc}_k} + X\beta, \tag{3}$$

$$\theta_j \sim \mathcal{N}(\tau, \sigma^2), \text{where } j = 1, \ldots, K_{loc}. \tag{4}$$

The parameters of the model are p_k, the true prevalence; θ, location-specific (random) effects on IFR; β, (fixed) effect of age; τ and σ, the hyper-mean and hyper-scale parameters for IFR. We implement our model in Stan[9], with weakly informative priors on all parameters. The prior for τ is centered at 1 death per 10,000 cases.

[3] The lower-risk group is defined as non-obese, non-smoking, and without the following risk factors (same as used by [43]): asthma, other chronic respiratory disease, chronic heart disease, diabetes mellitus, chronic liver disease, chronic neurological diseases, common autoimmune diseases (Rheumatoid Arthritis (RA), Systemic Lupus Erythematosus (SLE) or psoriasis), solid organ transplant, asplenia, other immunosuppressive conditions, cancer, evidence of reduced kidney function, and raised blood pressure or a diagnosis of hypertension.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

2.2 | Transforming individual risk into study risk

Once a suitable challenge virus is manufactured, itself a complex process [10], the risk of the individual from a challenge trial depends on the dose of virus given. The uncertainties about dose-response lead to a number of additional uncertainties about overall study risk. For other viruses, such as H1N1 and H3N2 influenza, a dose-response relationship has been found[26, 15]. The specific dose-response relationship, and its functional form must be determined experimentally, which is an outcome rather than an input of a dosing study like the one we are considering[4]. This uncertainty is a key issue, so, as suggested by Morgan and Henrion[28], we advise that this structural uncertainty should not be treated as a probabilistic variable, and instead sensitivity analysis should be used to enable the consideration of a range of plausible outcomes.

Given the specifics of a study design, the relationship between individual risk and the risks in the overall study is straightforward, assuming independence of risk between individuals in the study[5]. The risk to an individual of severe disease in the study given dose d is indicated S_d, and the risk of mortality is correspondingly M_d, so the probability that someone in a group of size N experiences the corresponding outcome is $1 - (1 - S_d)^N$ and $1 - (1 - M_d)^N$. By simulating the probability of impact for each dosed group and, in the case of more complex studies, conditioning the trial of later groups on the results of earlier ones, we can find the overall risk in more complex studies.

In the current version we have restricted the tool to consider to a simple N-person study. This means the tool provides an upper bound for risk of mortality and hospitalisation, higher than in a more complex dose-escalation study[6] However, the underlying model structure allows simulation of dose-dependent and/or conditional study designs, as necessary, to accommodate dose response information and allow for the simulation of more complex trials.

2.3 | Interface

The web tool we built allows exploration of two related types of risk. The first allows individuals to explore their personal risk, for example, by gender and age, if they volunteer to participate in a study, while the second displays overall risk of a single round dosing study with a given number of participants. Assumptions on study design and the underlying risk can be adjusted to allow interactive exploration by policymakers and participants.

In considering personal risk, the tool uses pre-calculated outputs from the risk model to calculate and display predictions for hospitalization rate and death rate for individuals as a function of age and sex.

Deaths are further translated into micromorts–the expected number of deaths per million events, a standard method of showing mortality risk[16] often used for patient consent [3]. Quantities which cannot be transformed into micromorts, are presented as probabilities[7].

The tool allows both the public and policymakers alike to explore how overall risks change depending on differences in study design. This also helps maintain transparency into clinical trial design concerns, thereby better informing potential challenge trial volunteers. We also note that for trial designers and ethicists, the relationship between risk of different impacts and compensation is critical[14, 33, 5].

The importance of transparency and public engagement has been widely noted in the literature on challenge trials[17]. For this reason, the tool is available for public use, and is already being used to inform people who have

[4] There is an assumption implicit in the model that the trial uses dose escalation or other data to ensure that the dose given does not greatly exceed the typical natural dose, or that a larger-than natural dose does not increase disease severity.

[5] The assumed independence is conditional on the age and health status of participants, and for dose-response studies, also infection severity by dose.

[6] This is equivalent to assuming that the risk of infection at each dose is above the threshold for infection to replace a dose-response curve.

[7] See discussion below about incorporating other risks in the model as more data becomes available.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

volunteered to be contacted to potentially participate in a challenge trial[1].

3 | RESULTS

The analysis dataset for age-specific IFRs contains 114 data points from 23 studies, with each containing between 3 and 11 different age groups; all data are presented in Table 2 in Appendix and are included in the code repository accompanying this paper. A glance at available data also confirms the necessity of using a more complicated modelling approach: out of 114 age-specific estimates, only 24 contain individuals aged between 20 and 30. However, out of these 24, only 2 have median age that falls between 20 and 30.

The basic evidence synthesis model, which includes all data, but does not adjust for health status, finds that average IFR in 20-29 age group for the studies included in this analysis is 17.5 per 100,000 cases (95% uncertainty interval[8] 12.7 to 23.3). Extending the HCT population to also include 30-39 year olds would lead to mean IFR of 29.9 per 100,000 (95% interval 21.9 to 39.9)[9].

Based on OpenSAFELY data, we estimate that in healthy population (defined as lack of any co-morbidities listed above), the average mortality risk in 20-29 year olds[10] is 1.9 times lower than in the general population, with 95% interval from 1.3 to 2.8.

We also note that there is large heterogeneity across the studies, due to treatment availability and other factors[11]. Because the HCT volunteer population will receive the best available care, including the most up-to-date treatment options, we can consider our population akin to a best-case scenario. In Belgium, the estimated risk for 20-29 year olds is 5.74×10^{-5}, approximately $\frac{1}{3}$ the estimated risk across all populations. Expanding the study population to 20-39 year olds would result in mean IFR of 9.88 per 100,000 cases.

Additionally, in the meta-analysis we estimate a population-level IFR by age, which does not account for the health status of screened volunteers. Looking at the OpenSAFELY dataset for the UK, we find a mean risk reduction factor in healthy populations aged 20-29 equal to 1.9, although our estimate is uncertain, with 95% interval from 1.3 to 2.8. This adjustment can be used to modify the risk estimate for the general population across studies.

It is unclear how much double-counting occurs when adjusting for both available medical care and for screening out comorbidities. However, using this estimate, we use our estimate that a simple 50-person challenge trial has a 99.85% probability of having no induced fatalities, and a 98.6% probability of having no cases serious enough to require hospitalization.

Due to lack of reliable data, the current model does not include estimates of longer term impacts, and instead includes a more qualitative discussion of these risks[38], but the model and interface will be updated with such estimates as they become available.

The model interface can be used to explore how these uncertain factors can interact, as shown below in Figure 2.

The overall study risk for a simple 50 person study is 0.15% probability of any deaths, and 1.4% probability of at least 1 hospitalization. This represents an upper bound of risk for 3 groups 10 of volunteers given different doses of COVID-19, and an additional 20 person cohort to ensure sufficient sample size.

[8] All intervals reported here are Bayesian posterior intervals. For brevity we just refer to them as "x% interval".

[9] See estimates in the appendix for how relative risk was estimated between age groups.

[10] Expanding to 20-39 year olds only slightly increases the risk reduction factor, from 1.9 to 2.

[11] Denoting by σ the hyper-scale parameter in the hierarchical model, 2σ impact corresponds to 3.96-fold mean decrease in IFR. That means we expect 2.5% of studies to have IFR more than 4 times lower than the average IFR.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

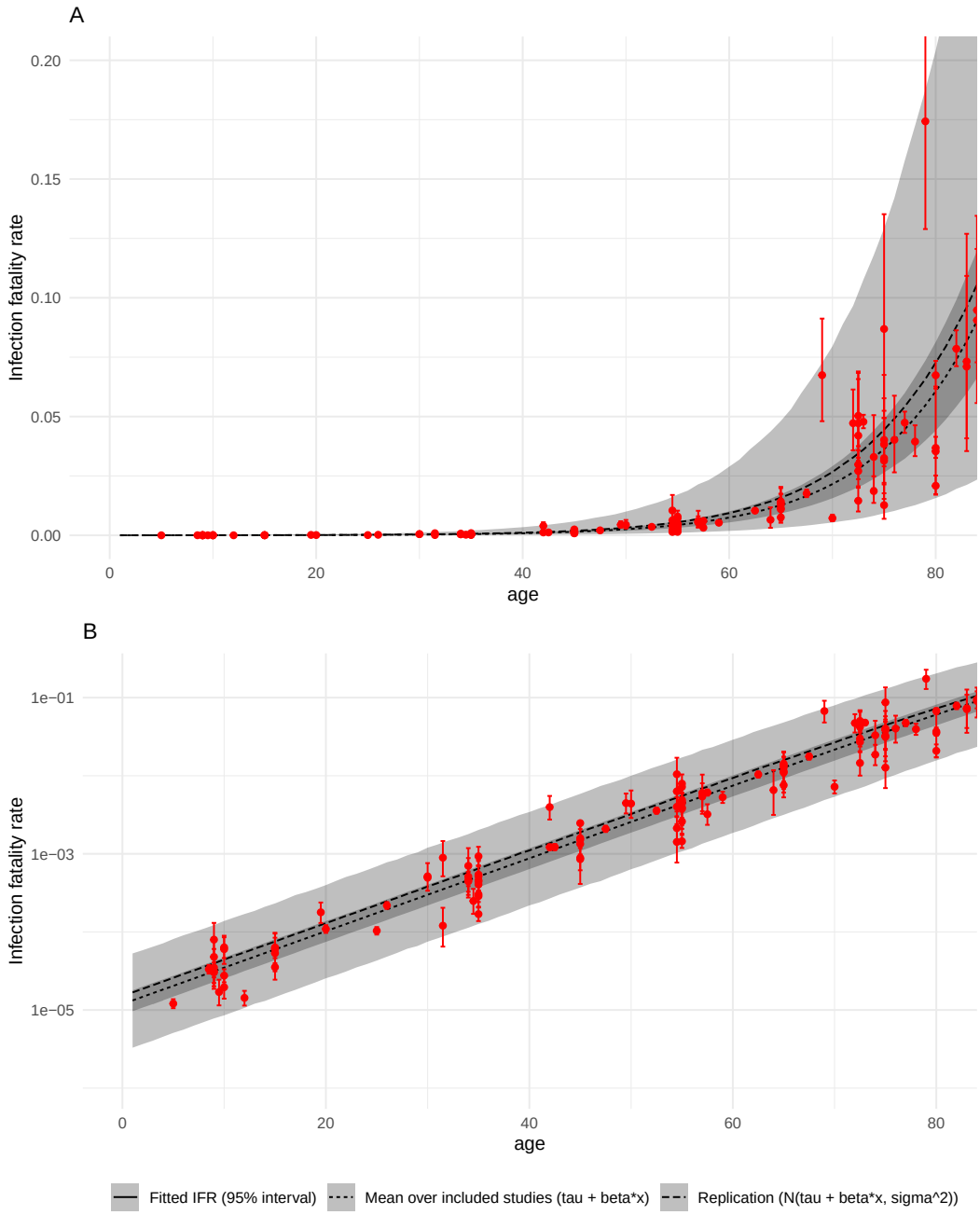

FIGURE 1 IFR as a function of age. Panel A is untransformed data. Panel B shows the same data on log 10 scale. Red points are data, *i.e.* model estimates of mean IFRs in particular studies, with bars representing 95% posterior intervals. Black and gray are modelled IFRs: lines are means, ribbons are 95% posteriors intervals: the narrower is average across all included studies, the wider takes into account heterogeneity between studies. Details and input values are given in Appendix.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

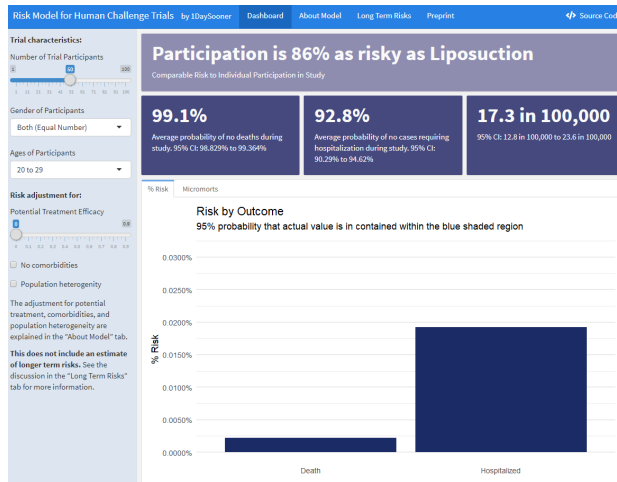

FIGURE 2 Interactive Model Interface

While the estimate is specific to a dosing study, it can also be useful for understanding the risk of later vaccine trials, though in that case other factors, including the possibility of vaccine-enhanced infection, would need to be assessed.

4 | DISCUSSION

We have demonstrated that we can model certain risks associated with human challenge trials. However, our model is incomplete, and considerable concerns remain about the risks of human challenge trials. While a full accounting of challenge trial ethics is beyond the scope of this paper, we consider several factors below that inform our work on human challenge trial design, especially unmodeled risks. For a more complete perspective on the ethics of COVID-19 HCTs, we direct the reader to the World Health Organizations' (WHO) key considerations [31] and the recent discussion in Lancet Infectious Disease [17, 24] on how these issues are being addressed.

In our modal case, it is clear that the risk of our dosage-response study is not higher (and indeed likely far lower) than the risk from comparable clinical infections in the general population, and are far lower than other risks that are typically widely viewed as acceptable. Further, human challenge trials have had historical precedent, showing promise for both less lethal human coronaviruses, and for yellow fever[37]. They also provided early indications regarding the possible efficacy of a leading malaria vaccine candidate[30]. The discussions about earlier trials shows that the ethical challenges of HCTs should not be seen as unique, but rather as lying along a natural continuum of clinical studies [12]. That said, the initial human challenge trials should be held to high ethical standards, both for individual risks and to preserve public trust in scientific and medical progress. That includes emphasis on fully informed consent of the participants, and as noted by the WHO, higher than typical ethical standards[31].

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

4.1 | Opportunity Cost

When evaluating clinical trial design, it is not sufficient to evaluate whether the proposed model is good *ex nihilo*. One must consider whether the alternatives are better. The key benefit of a dosing study is to allow further research with challenge trials, and alternative clinical trial models have major practical difficulties and far higher costs in a variety of ways. While this is not relevant for the initial set of vaccines that are close to approval, the challenge of large scale trials is magnified for later vaccines.

For example, "standard" phase 3 efficacy trials for an ongoing novel pandemic rely on high numbers of trial participants, and require this large sample for each vaccine or treatment that is trialled. Such trials are relatively expensive, and expose more trial participants to negative side effects of a given treatment, so that in many scenarios HCTs have been shown to be superior [4]. Standard trials are also difficult to pursue after an initial vaccine is available, and in the likely and hoped-for case that there will soon be at least one vaccine, finding willing participants for later and perhaps more effective vaccines, or ones with fewer side effects, is harder, and the burden of proof for benefit is far higher, or not achievable given plausible sample sizes, without an HCT.

Assuming a log-linear relationship between age and IFR, we estimated that IFR increases on average 2.93-fold per each additional decade of age. We also estimate IFR of 1.75 per 10,000 among individuals aged 20-30, with considerable heterogeneity in IFR across locations, ranging from 6.29 per 10,000 to 5.7 per 100,000. These results align with two recent meta-analyses of IFRs by O'Driscoll *et al* and by Brazeau *et al* [32] [6]. Both studies also show log-linear relationships between age and IFR, as well as considerable variation in IFR by context. O'Driscoll *et al* report median IFRs of 0.6 and 1.3 per 10,000 among those aged 20-24 and 25-29, respectively, while Brazeau *et al* report 3 and 4 per 10,000 respectively.

4.2 | Limitations

Finally, it is important to note that our model does have limitations. Hospitalization rates and death for 20-30 year olds are rare; our prior knowledge on fatalities for this group are more uncertain due to limited data. Information on long-term damage caused by COVID-19 is similarly incomplete, and though this is discussed further below, our model does not currently account for that risk. Also note that although our model uses hospitalization as a proxy for the upper bound of serious nonfatal COVID-19 cases, more data is required to see if this is an accurate assumption.

Finally, our model may not accurately capture changes in COVID-19 risks over time. It also does not estimate any indirect risks of the study. We stress that our model is not a comprehensive analysis of all available risks, but rather a tool quantifying certain known risks that can be used by trial participants and policymakers.

4.3 | Non-modeled Risks

We also note that there are several impacts we do not model in the study, most notably the concern about so-called Long-COVID, which is a catch-all term referring to a combination of persistent symptoms, i.e. slow recovery, and new post-recovery symptoms[7]. It is understood that for some cases, especially severe ones, recovery from COVID-19 can take months. In others, there are longer-term symptoms differing from those experienced during the infection, perhaps similar to Post-SARS syndrome[34, 27]. At the same time, COVID-19 recovery has been found to be faster in younger, healthier patients[40], which may mean the risk is lower in this group. It seems clear that the risk is a subject of continuing scientific investigation, and as it becomes better understood and quantified, it will be incorporated into both the risk model, and the web tool. Until then, the model contains an embedded overview of what is currently

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

understood and known or unknown about longer term risks[38].

We also note that vaccine-induced disease enhancement is a critical concern for vaccine challenge trials, but is not relevant to dosing studies. Still, this risk must be considered in analyses of risks for later trials, and the model would need to be adapted or supplemented to consider this.

5 | CONCLUSION

Human challenge trials are not risk free, but the balance of risk and benefits seems to clearly favor allowing them, as a large group of experts has argued[39]. This conclusion is disputed by some, but all decisions are made despite uncertainties and debate - whether empirical or moral[22, 23]. The question is whether the both empirical and moral balance of factors lead to the indicated conclusion. The alternative is a failure to act due to misguided risk-aversion, or worse, using uncertainty and disputed moral claims as a positive stance to shut down further work, as has occurred in the debate about HCTs[25].

It seems likely that Challenge trials are a viable way to rapidly test vaccine efficacy, which is particularly critical now for testing second-generation vaccines, which may prove superior to first generation vaccines, or at least help fill the demand unmet by first-generation candidates. A dosing study is an urgent first step, and the risk estimates and tools developed for this paper can assist in planning such studies and informing volunteers.

Our model provides insight into the overall risk of a trial of a given size, and can better inform HCT participants about the dangers they face. Given that an HCT may help select the multiple vaccines necessary for global immunization while also assisting with therapeutic testing, the risk of an initial study into SARS-CoV-2 pathogenesis seems justified. If the dosing study is successful, future HCTs of COVID-19 may provide a rapid and systematic way of screening vaccine candidates for efficacy and safety, which is a significant benefit.

The results presented here are a useful static estimate of risk. The model is already being used to inform potential volunteers [1], and can be adapted and expanded in the future. Given the evolving understanding of the disease, the model should be continually updated with additional data on mortality and hospitalisations risks or other long-term risks. This will contribute to the discussion of whether or not to pursue challenge trials, which can help with response to COVID-19 in a variety of ways[29]. Challenge trials may be an important tool for fighting COVID-19, and our model is a step towards that goal.

Acknowledgements: Thanks to Chris Choe, Troy Yamaguchi, and Linchuan Zhang for help preparing and editing the manuscript.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Appendix: Bayesian model of COVID-19 mortality risk in HCT volunteers

A-1 | INTRODUCTION

This short document is a technical appendix to the paper discussing COVID-19 risks in human challenge trial. Here, we show how a Bayesian model can synthesise information on many infection fatality rates (IFRs) into a single estimate. This estimate is specific to certain age groups and can be further adjusted by e.g. co-morbidity status. The analysis presented here is a form of Bayesian meta-analysis, and our primary objective is to weigh sources of evidence in a way that captures both variability (here, heterogeneity in real IFRs across different settings) and uncertainty (here, the fact that we do not know the IFRs in each setting precisely).

The ultimate objective of this model is to characterise risk in a way that is useful for design of HCTs. Therefore, as a minimum, we want to incorporate variability across different populations into our prediction. Even better would be to understand how different factors can drive heterogeneity: *a priori* we hypothesise that the three main drivers of differences in IFRs are time-specific, population-specific and otherwise country-specific[1].

To characterise differences in observed IFRs we first develop a Bayesian model and apply it to publicly available summary data on IFRs from multiple countries and contexts, with particular focus on the impact of age. This is covered by Section A-2. We then use a simple model to hypothesise reduction in risk that may be achieved by screening individuals for comorbidities; this is Section A-2.2. We summarise all results in Section A-2.3.

A-2 | AGE-SPECIFIC RISK OF COVID-19 MORTALITY

A-2.1 | Bayesian evidence synthesis model

What follows is an adaptation of typical methods of Bayesian evidence synthesis to analysis of IFRs. IFR is the ratio of deaths to infections in a given population. Early estimates of Covid-19 mortality risk, e.g. by Verity et al. [42], placed it at over 0.6%; however, it was also evident from data that IFR could be orders of magnitude higher in particular high risk groups, especially in the elderly, than in the general population[2].

By definition, our data on IFRs is a combination of data on deaths with data on infections. Typically, these are disjoint samples, in that numbers of infections are estimated (typically very imprecisely) on select subpopulation, while deaths are recorded in the general population (at a level of country, administrative region etc.). There are clear reasons to believe that IFRs will differ across studies (e.g. due to age, comorbidity status, time, genetic factors, quality of healthcare etc.). To address this, we will use a Bayesian hierarchical modelling framework to assume that the setting-specific estimates of IFR_k can differ from each other but are linked through some common parameters. (By k's we denote different populations; note that sometimes we may have multiple IFR's from different age groups in the same location.)

The most straight-forward and "canonical" way to implement such a Bayesian model is by modelling log odds of the event[3]. Deeks [11] present a general treatment of such approach in medical statistics. Note, that for very rare

[1] The role of time may be due to new treatments, improvements over time in our ability to treat COVID-19 or selection pressures which may lead to more benign versions of the virus. Country-specific or location-specific factors in IFR data may be driven by under-reporting, health care factors (including access to health care services) or underlying distributions of known risk factors. Additionally, some unknown risk factors (e.g. genetic) may also be operating, in which case controlling for age and co-morbidities will be not sufficient to account for cross-location differences.

[2] Various estimates published since suggested that the relationship of mortality risk to age is consistent across different countries.

[3] It is also possible to work with IFR_k parameters and treat them as derived from Beta distribution with some "hyperparameters" α and β

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

events the odds of mortality are very similar to probability of mortality, but we model events on odds scale as a good "generic" approach to modelling binary data (in this case death following infections)[4].

Basic models for this type of analysis of binary data can be implemented using existing statistical analysis packages (see, for example, *metafor* package in R or *baggr* by Więcek and Meager [44]), by treating IFR as a logit-normal parameter to meta-analyse. However, note that when no deaths are observed, analysis of IFR (equal to observed deaths divided by modelled infections) is problematic. Therefore we propose a "custom" model that built in Stan which treats deaths and *prevalences* (rather than the IFRs) as data.

Let d_k denote observed deaths for data point k and assume that logit of corresponding prevalence estimate is p_k is a parameter (typically obtained from a statistical modelling papers, government reports etc.). Total population in k-th setting is n_k. Total number of estimates is K. Then the model likelihood is as follows:

$$d_k \sim \text{Binomial}(n_k, p_k IFR_k) \tag{5}$$

$$\text{logit}(p_k) \sim \mathcal{N}(\mu_k^{(p)}, (\sigma_k^{(p)})^2) \tag{6}$$

where $\sigma_k^{(p)}$ and $\mu_k^{(p)}$ are parameters obtained from the literature (or converted from these parameters – see next section). The k data points collected can span many locations (studies); we denote them by loc_k and the total number of locations by K_{loc} (with $K_{loc} < K$).

In this model we can also account for various covariates impacting the IFRs (let's denote their total number by N_p), such as age groups (which we identify with median age of the population being studied, MedianAge_k). We code them in a design matrix X. To center our X at the value of interest in our model (risk in 20-30 year olds), we use a transformation `MedianAge/10 - 2.5` to construct our matrix X. We denote all of the covariates using a design matrix X and denote by N_p the number of columns in X. We assume the impact on IFR is on logit scale, same as in the "canonical" logistic models of binary data that we mentioned above:

$$\text{logit}(IFR_k) = \theta_{\text{loc}_k} + X\beta, \tag{7}$$

$$\theta_j \sim \mathcal{N}(\tau, \sigma^2), \text{where } j = 1, \ldots, K_{loc}. \tag{8}$$

This means θ spans location-specific (random) effects on IFR while β is N_p dimensional vector of (fixed) covariate effects.

We implement our model in Stan and assume weakly informative priors on all parameters, with prior for τ centered at 1 death per 10,000 cases.

```
model {
  //Uncertain prevalence estimates (mu^p and sigma^p above):
  logit_prevalence ~ normal(mean_prevalence, sd_prevalence);
  //Likelihood of mortality (d_k = obs_deaths, p_k = prevalence):
```

of Beta distribution, as done by e.g. Carpenter[8]. That approach, however, does not offer an easy way of modelling impact of covariates (e.g. age and co-morbidities) on the rates.

[4] Another advantage of such a model is that it can use either individual-level or summary data and work with covariates (such as gender, age, time of the study, co-morbidities), captured as odds ratios or risk ratios. If only summary data are available, covariates can be defined as study level distributions (e.g. % male).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .


```
  obs_deaths ~ binomial(population, prevalence .* ifr);
  //Hierarchical component of the model (location-specific theta):
  //logit_ifr = theta_k[loc] + to_vector(X*beta);
  theta_k ~ normal(tau, sigma);

  //Priors:
  tau   ~ normal(logit(.0001), 5);
  sigma ~ normal(0, 10);
  beta  ~ normal(0, 10);
}
```

A-2.2 | Data

We used estimates originally collected by Levin, Cochran, and Walsh (2020) to construct the first version of analysis dataset, which we then supplemented with more values extracted from other studies. The input data into our model consists of deaths (treated as known) and prevalences (treated as logit-distributed parameter with known mean and SD) in all reported age groups in all studies[5].

All of input data are given in Table 2. The analysis dataset contains 114 data points from 23 studies, each containing between 3 and 11 different age groups. We made only minimal modifications to source data, by 1) imputing the values in the Italian fatality data based on a seroprevalence survey, 2) imputing population size in Maranhao (as ratio of the reported number of infections and the mean infection rate) which were not reported and 3) assuming that uncertainty in prevalence 0-29 age group in Iceland is same as in the 30-39 age group since data were missing.

As mentioned, our model treats number of COVID-attributable deaths as measured without error (due to lack of data) but accounts for uncertainty in infection rates, which are always model-based estimates extracted from various available data sources. In preparing data, we assumed that logits of prevalence estimates from available studies are normally distributed, which seems to reproduce majority of data very well, see Figure 7 in the Supplement. There are some discrepancies with studies that allowed for prevalence estimates to be 0, something that our logit model does not allow.

For each study we construct a median age scalar defined by the average of the endpoints of each age range, rather than attmepting to calculate a population-weighted mean.

Our approach of regressing on the median age and use of all available data (rather than the subset of data available in younger adults only) is necessitated by data limitations: out of 114 data points comprising age-specific estimates of prevalence (or IFR) and counts of deaths, 24 contain individuals aged 20-30 who are of primary interest to us. However, the populations are mixed with regards to age, with typical age groupings such as 19-49, 20-49, 20-39, 0-49 used instead. In fact, we find only one estimate out of 24 that is entirely specific to the 20-29 age group (Brazilian state of Maranhao), while one more has median age falling between 20 and 30 but is not specific to that age group.

A-2.3 | Results

There were no issues with convergence of the Bayesian model. We set number of iterations to 5,000 and used 4 chains, with `max_treedepth` option set to 15. There were no divergent transitions and effective sample size was

[5] This basic approach exaggerates uncertainty, as we treat different 95% intervals reported in the study as uncorrelated.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

greater than 3190 for all of 368 modeled parameters (this number includes fitted prevalences, IFRs, θ's and their transformations into/from logit scales). For the three main parameters in the model we obtained the following:

```
## Inference for Stan model: ifr_with0.
## 4 chains, each with iter=5000; warmup=2500; thin=1;
## post-warmup draws per chain=2500, total post-warmup draws=10000.
##
##          mean se_mean   sd  2.5%   25%   50%   75%   98% n_eff Rhat
## tau     -8.66       0 0.15 -8.97 -8.76 -8.66 -8.56 -8.35 10140    1
## sigma    0.67       0 0.12  0.48  0.59  0.66  0.74  0.96  9711    1
## beta[1]  1.07       0 0.01  1.05  1.07  1.07  1.08  1.10  4217    1
##
## Samples were drawn using NUTS(diag_e) at Mon Nov 02 09:51:57 2020.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

The mean coefficient of beta 1.07 corresponds to 2.93-fold increase in mortality risk following an infection per each extra decade of age (95% uncertainty interval is 2.87-2.99).

From these parameters we can predict average risks for subjects of any given age x, by using the posterior distribution of $\tau + (10x + 2.5)\beta$ (where 2.5 and 10 refer to the transformation that we applied to `MedianAge` inputs).

A-2.4 | Average infection fatality risk in young subjects

Since we centered our `MedianAge` at 25 years in constructing our matrix X, we can now obtain model-estimated risk for a typical HCT population (aged 20 to 30, with median 25) by ignoring the β coefficient and examining τ and σ only. We find that the average IFR for this group (equal to $\frac{\exp(\tau)}{\exp(1+\tau)}$) is 1.75×10^{-4} (with 95% interval from 1.28×10^{-4} to 2.36×10^{-4}). That means, on average, slightly under 2 deaths per 10,000 infections in the studied datasets.

A-2.4.1 | Heterogeneity in IFRs

However, there is a considerable variability in IFRs across different locations/dataset that we should consider. To take into account parameter σ, we can generate draws from the $\mathcal{N}(\tau, \sigma^2)$ distribution, corresponding to a hypothetical IFR in a new source of data. 95% interval for such model runs from 4.39×10^{-5} to 6.94×10^{-4}. Since the model works a logistic scale, another way of interpreting the across-dataset variability is reporting the fold-impact of σ on the mean IFR; here, we obtain on average a 3.96-fold increase (decrease) in IFR per 2σ increase (decrease).

The lower end of the 95% interval, 4.39×10^{-5}, is not extreme given input data, where the "crude" mean IFR (based on mean prevalence only) is below 7 per 10,000 for all data except for South Florida, and as low as 0 for some countries that did not record deaths (Belgium, New Zealand, Korea, Iceland) in various age groups including 20-29 year olds or 1.4 per 10,000 in Utah, in the population aged 19-44. (Please refer to Table 2 for complete list of inputs.)

We can assess this heterogeneity by inspecting the distribution of random effects in the model transformed into IFRs, i.e. the inverse logit transformation θ parameters. The largest (posterior mean) IFR value of θ is 6.3×10^{-4} in Castiglione d'Adda. The smallest posterior mean for 20-29 year olds is 5.73×10^{-5} in Utah.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

A-2.4.2 | Predictive checks for the model

We constructed posterior predictive distributions for number of deaths in each of the inputs by using the `generated quantities` functionality of Stan. Figure 3 compares the posterior means and 95% intervals with observed deaths. Out of 114 observations that were used to fit the model, 109 were within 95% intervals of the posterior predictive distributions. We observed the largest discrepancies occurred in Spanish data. Overall, we conclude that the simple binomial model we used here is flexible enough to capture both age-specific risk increases and heterogeneity in IFRs across settings/countries.

A-2.5 | Sensitivity analyses

Planned sensitivity analyses include:

- Exclusion of small studies and test & trace data
- Median age: exclusion of 80+ population (non-linear behaviour on logit IFR?)
  - Could also check exclusion of children and teenagers, due to imprecise prevalence numbers
- Median age: different method of age imputation
- Model including time

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

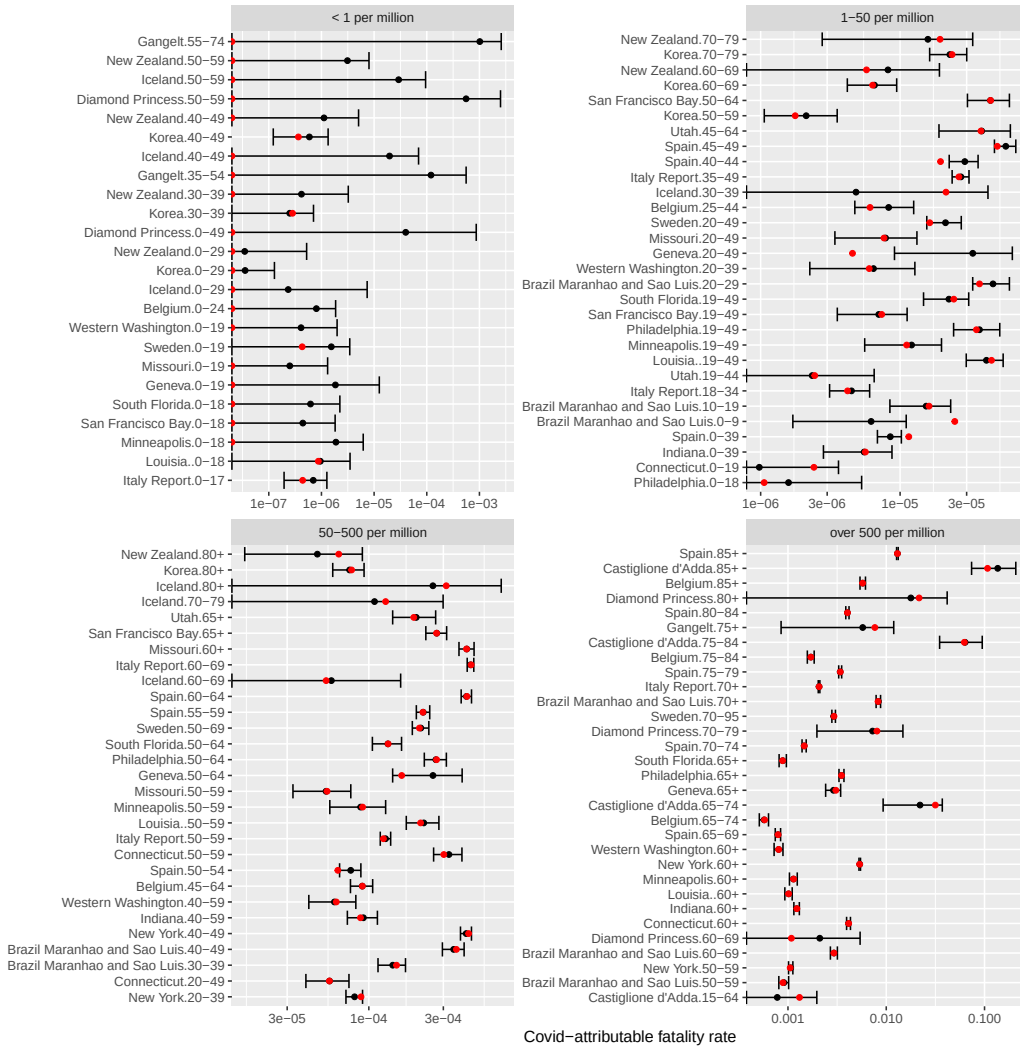

FIGURE 3 Comparison of model estimates (black) with data on observed fatality rate (FR, red), compared on logarithmic scale. FR is number of deaths divided by overall population size. Bars are 95% posterior interval; point is the mean. For better clarity, we grouped the plot into four panels according to observed FR X axes on each panel differ. For many low-risk populations (upper-left quadrant) no deaths were reported: we indicate this by plotting a red point on the left-hand side of the panel plot.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

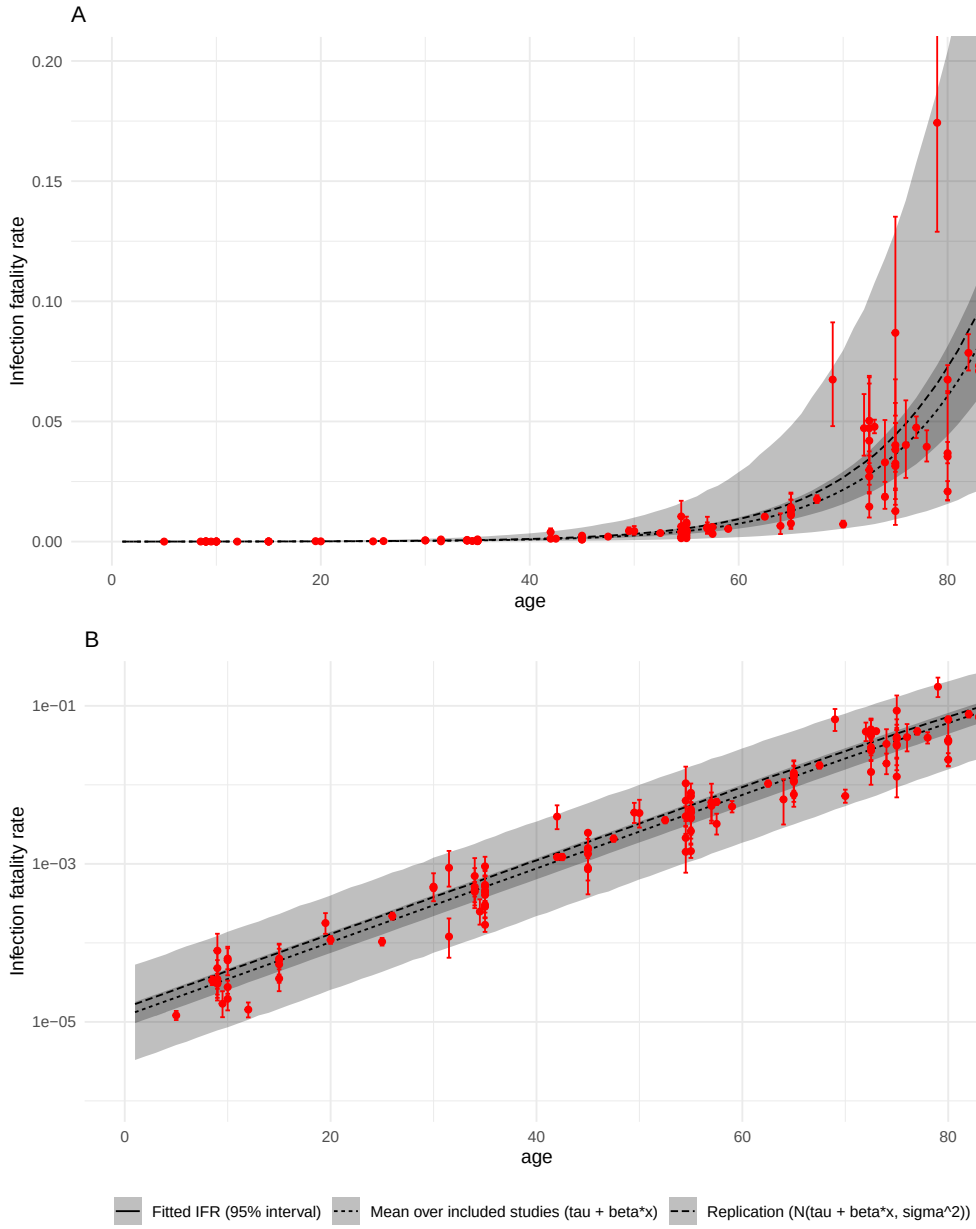

FIGURE 4 IFR as a function of age. Narrower ribbon corresponds to the 95% posterior interval of average across all included studies (tau parameter in the meta-analysis model), while the wider band takes into account heterogeneity (tau and sigma). Lines are means. Red points are model estimates of mean IFRs in partciular studies, with bars representing 95% posterior intervals. Panel A is untransformed data. Panel B shows the same data on log 10 scale.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

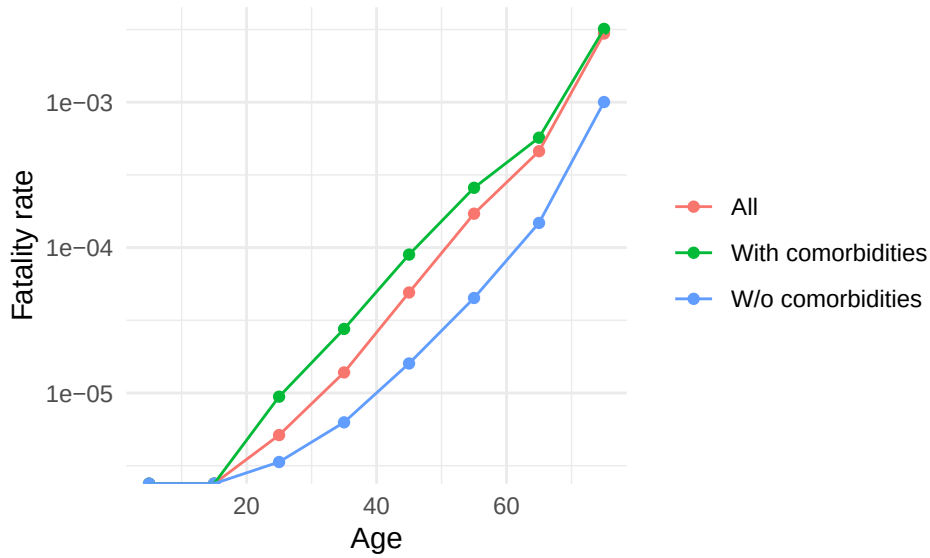

FIGURE 5 Fatality rates as a function of age in OpenSAFELY data. Fatality rate is zero in the 0-10 and 10-20 age groups.

A-3 | RISK REDUCTION IN HEALTHY INDIVIDUALS

We now turn our attention to the question of how much a human challenge trial designer could reduce the mortality risk by using simple screening methods, as discussed in the main paper.

Data for this section has been provided by OpenSAFELY (`https://opensafely.org/`) and was used by Williamson et al. [43] to characterise COVID-19 mortality risk factors for 10,926 COVID-19 deaths in England. We group the total of 21,444,863 individuals into a total population and a lower-risk sub-population, defined as non-smoker, non-obese and without the comoribidities reported in the OpenSAFELY study[6], most notably respiratory and cardiovascular diseases and type I diabetes. For brevity we refer to the population without one of the pre-defined comorbidities as "healthy". In contrast to the cited publication, we include records of individuals under 18 in our assessment. Counts grouped by age are presented in Table 1.

As shown in Table 1, for the age group of 20-29 the crude risk ratio (of general population vs the healthy subset only) is 1.53, but, due to low number of events in both healthy and general population, with a very wide 95% interval from 0.6 to 5.65[7]. As data on relative risks in other age groups is clearly related to the relative risk in 20-29 age group, we use another meta-analysis model to improve our estimate. Additionally, relative risks are higher in women than in men – something that we can account for in our model too.

In our modelling we make a strong assumption that infection rates in population with comorbidities are the same as in the general population. In other words, we assume that denominator for IFR is same in both populations. We

[6] "asthma, other chronic respiratory disease, chronic heart disease, diabetes mellitus, chronic liver disease, chronic neurological diseases, common autoimmune diseases (Rheumatoid Arthritis (RA), Systemic Lupus Erythematosus (SLE) or psoriasis), solid organ transplant, asplenia, other immunosuppressive conditions, cancer, evidence of reduced kidney function, and raised blood pressure or a diagnosis of hypertension"

[7] We obtain the interval using a simulation approach. Using normal approximation of log(RR) statistic we obtain a narrower 0.57 to 4.1, perhaps due to poor quality of approximation for rare events.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE 1 Data from the OpenSAFELY database grouped by age.

	Population		Fatalities		FR per 100k		
Age group	All	Healthy	All	Healthy	All	Healthy	RR
0 to 9	2,160,958	2,007,997	0	0	0.00	0.00	NaN
10 to 19	2,428,494	2,000,761	0	0	0.00	0.00	NaN
20 to 29	2,530,792	1,788,907	13	6	0.51	0.34	1.5
30 to 39	2,960,611	1,909,227	41	12	1.38	0.63	2.2
40 to 49	2,849,984	1,565,935	140	25	4.91	1.60	3.1
50 to 59	3,051,110	1,243,728	522	56	17.11	4.50	3.8
60 to 69	2,392,392	622,357	1,101	92	46.02	14.78	3.1
70 to Inf	3,070,522	317,238	9,109	318	296.66	100.24	3.0

then specify a generic partial pooling model of fatality rates (FR, defined as number of deaths in the entire population, without regards to infection status) such that

$$\text{logit}(\text{FR}_k) \sim \mathcal{N}(\alpha_{\text{age}_k}\text{comorb}_k + \beta\text{male}_k + \gamma_{\text{age}_k}), \tag{9}$$

$$\alpha_i \sim \mathcal{N}(\mu, \sigma) \text{ for all } i, \tag{10}$$

where, for k-th observation, age_k is the age group, and comorb_k and male_k are indicator variables. This means that each age group is assigned different "baseline" fatality rates.

Note that the assumption of FRs varying across age groups that we just mentioned differs from the model of age-specific IFRs in Section 2. This is because we hypothesised that infection rates (which are used as denominators in the IFR model of Section 2) will vary across age groups. This is borne out by Figure 6.

Summary of the main model parameters is as follows:

```
##            mean se_mean    sd   2.5%    25%    50%    75%    98% n_eff Rhat
## mu         1.43 0.00277 0.192   1.07   1.33   1.43   1.54   1.83  4810    1
## sigma      0.36 0.00596 0.226   0.12   0.22   0.31   0.44   0.93  1443    1
## alpha[3]   1.32 0.00429 0.328   0.61   1.14   1.34   1.53   1.94  5859    1
## beta       0.40 0.00019 0.019   0.37   0.39   0.40   0.41   0.44  9969    1
## gamma[3] -13.06 0.00404 0.344 -13.77 -13.28 -13.04 -12.82 -12.42  7227    1
```

We find that the mean risk ratio between population with comorbidities and healthy sub-population in 20-29 age group (exponent of α_3 above) is 3.96, with wide 95% uncertainty interval of 1.85 to 6.96[8].

[8] Using simple models that assumed identical risk ratios in all age groups would lead to a mean RR of similar magnitude but a much less uncertain estimate, due to more rigid model assumptions; similarly, assuming some linear age structure on risks, such as in the main meta-analysis model above, would may lead to a different RR, but we do not think such an assumption is justified here. We do not include outputs of these models in this short write-up.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

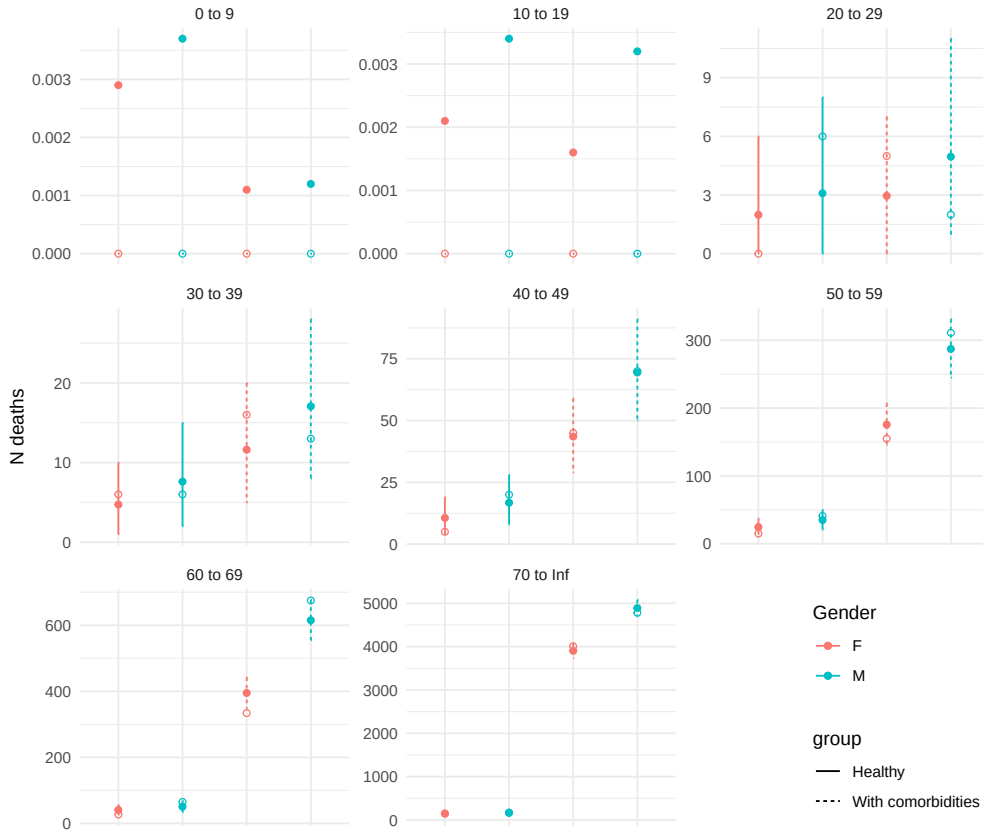

FIGURE 6 Comparison of posterior predictive numbers of deaths from the fitted Bayesian model (mean and 95% uncertainty intervals) with data inputs (circles). For each age grouping we have 4 estimates: male/female and healthy vs general population

Next, using the posterior samples we calculate the event rate in total population (i.e. $\theta* = (\theta_1 n_1 + \theta_2 n_2)/(n_1 + n_2)$, where subscripts 1 and 2 are healthy and comorbid sub-populations) and then divide it by ratio in healthy population to obtain an estimate of risk reduction possible by selecting healthy volunteers only. The mean posterior value is 1.88, with 95% uncertainty interval from 1.25 to 2.76. Due to use of Bayesian hierarchical model over many age groups, the uncertainty interval is much narrower than on the risk reduction factor calculated on 20-29 age group only.

To validate the model, we conducted a simple posterior predictive check for numbers of deaths in different age groups and genders. The graphical check is presented in Figure 6. Overall we find that the simple model has no problem with reproducing observed data.

A-4 | CONCLUSION AND SUMMARY OF RESULTS

In conclusion, the implications of the model for the risk in healthy young subjects are as follows:

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

- We find that average IFR in 20-29 age group for the studies included in this analysis is 1.75×10^{-4} with 95% interval from 1.28×10^{-4} to 2.36×10^{-4}.
  - It is feasible that the mean IFR can be decreased as much as 3.96-fold (2σ impact on the IFR according to hyper-SD parameter in the meta-analysis model).
  - It is easy to argue that a HCT designer would be able to achieve IFR at least as low as within any of the large-scale studies included in our sample of populations. The smallest posterior mean for 20-29 year olds is 5.73×10^{-5}, fitted to data from Utah.
  - Extending the HCT population to also include 30-39 year olds would lead to mean IFR of 2.99×10^{-4} with 95% interval from 2.19×10^{-4} to 4.02×10^{-4}. Lowest mean IFR would then be 1.02×10^{-4} (also in Utah).
- In healthy population (defined as lack of co-morbidities listed above), the average mortality risk in 20-29 year olds is 1.88 times lower than in the general population, with 95% uncertainty interval from 1.25 to 2.76.
  - Our 1.88 estimate is a bit higher than the mean "crude" risk ratio of 1.53 because we use a Bayesian hierarchical model that synthesises evidence across all age groups.
  - Expanding to 20-39 year olds, the risk in healthy sub-population would be 2.04 times lower than in the general population (95% interval from 1.25 to 2.76).
- Combining the smallest posterior IFR for 20-29 year olds with our estimated fold-reduction due to excluding individuals with co-morbidities from the population would lead to a mean infection fatality risk of 3.31×10^{-5} with 95% Bayesian interval from 1.53×10^{-5} to 6.13×10^{-5}.
  - In 20-39 year old subjects the risk would be 5.11×10^{-5} (2.54×10^{-5} to 9.04×10^{-5}).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

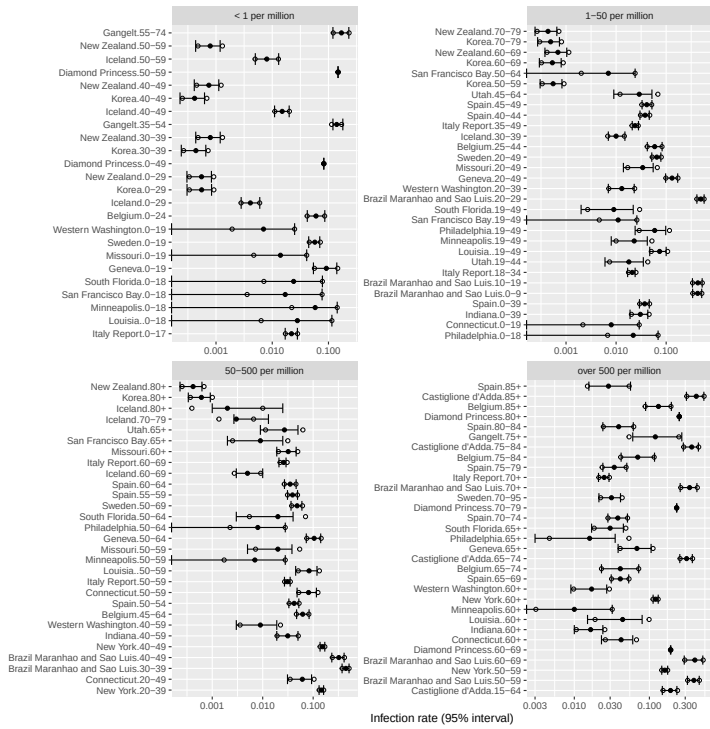

FIGURE 7 Comparison of model-estimated prevalences (95% CI's reported by modelling studies) collected by Levin, Cochran, and Walsh (2020) and our distributional assumptions: additional circles show 95% CIs recreated by assuming logit-normal distribution of prevalence. We group studies into 4 bands of mortality to mirror earlier figures. Please note that this approach produces discrepancies in a number of US estimates where the confidence intervals were skewed toward including 0. However, since we do not have access to source data, we decided to use the logit-normal assumption for all estimates. This assumption may have an effect of overestimating mortality risk in settings where prevalence was very low.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who
has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE 2 Complete table of inputs used by the meta-analysis model and crude IFR's.

							Infection rate (%)			
Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	0-9	29	1.2e+06	42.60	51.30	33.80	0.06
Italy Report	Country	Seroprevalence	2020-08-01	0-17	5	1.0e+07	2.20	2.80	1.70	0.02
Louisia.	USState	IFR	2020-05-06	0-18	1	1.1e+06	2.80	11.50	0.00	0.03
Minneapolis	USState	IFR	2020-06-04	0-18	0	9.7e+05	5.80	14.30	0.00	0.00
Philadelphia	USState	IFR	2020-05-17	0-18	1	9.4e+05	2.20	6.90	0.00	0.05
San Francisco Bay	USState	IFR	2020-05-25	0-18	0	1.6e+06	1.70	7.70	0.00	0.00
South Florida	USState	IFR	NA	0-18	0	1.3e+06	2.40	7.80	0.00	0.00
Missouri	USState	IFR	2020-05-23	0-19	0	1.5e+06	1.40	4.10	0.00	0.00
Connecticut	USState	IFR	2020-05-28	0-19	2	8.3e+05	0.80	2.90	0.00	0.30
Geneva	City	IFR	2020-06-01	0-19	0	8.0e+04	9.17	14.06	5.40	0.00
Sweden	Country	IFR	2020-06-18	0-19	1	2.3e+06	5.73	7.00	4.45	0.01
Western Washington	USState	IFR	NA	0-19	0	1.0e+06	0.70	2.50	0.00	0.00
Belgium	Country	IFR	2020-05-16	0-24	0	3.2e+06	6.00	8.60	4.20	0.00
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	10-19	21	1.3e+06	43.00	52.40	33.50	0.04
Iceland	Country	IFR	2020-06-14	0-29	0	1.4e+05	0.41	0.60	0.28	0.00
Korea	Country	IFR	2020-07-11	0-29	0	1.6e+07	0.06	0.08	0.03	0.00
New Zealand	Country	IFR	2020-07-09	0-29	0	1.9e+06	0.06	0.08	0.03	0.00
Indiana	USState	IFR	NA	0-39	20	3.5e+06	3.05	4.30	1.90	0.18
Spain	Country	IFR	2020-07-15	0-39	225	1.9e+07	3.73	4.60	2.93	0.31
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	20-29	47	1.3e+06	49.20	57.30	41.10	0.08
Italy Report	Country	Seroprevalence	2020-08-01	18-34	43	1.0e+07	2.10	2.40	1.70	0.20

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE 2 Complete table of inputs used by the meta-analysis model and crude IFR's. *(continued)*

Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
New York	USState	IFR	2020-05-23	20-39	482	5.4e+06	14.60	16.10	13.10	0.61
Western Washington	USState	IFR	NA	20-39	8	1.3e+06	1.30	2.30	0.70	0.46
Philadelphia	USState	IFR	2020-05-17	19-49	51	1.4e+06	5.90	9.80	2.40	0.60
Utah	City	IFR	2020-06-19	19-44	3	1.2e+06	1.80	3.50	0.60	0.14
Louisia.	USState	IFR	2020-05-06	19-49	85	1.9e+06	7.40	10.00	4.70	0.61
Minneapolis	USState	IFR	2020-06-04	19-49	18	1.6e+06	2.30	4.20	0.80	0.48
San Francisco Bay	USState	IFR	2020-05-25	19-49	25	3.4e+06	1.10	2.60	0.00	0.67
South Florida	USState	IFR	NA	19-49	61	2.5e+06	0.90	2.20	0.20	2.70
Missouri	USState	IFR	2020-05-23	20-49	18	2.3e+06	3.40	5.50	1.40	0.23
Belgium	Country	IFR	2020-05-16	25-44	18	3.0e+06	5.90	8.30	4.20	0.10
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	30-39	163	1.1e+06	44.40	51.40	37.40	0.34
Connecticut	USState	IFR	2020-05-28	20-49	75	1.3e+06	6.10	9.30	3.10	0.92
Diamond Princess	Small	IFR	2020-04-01	0-49	0	1.2e+03	8.26	8.28	8.24	0.00
Geneva	City	IFR	2020-06-01	20-49	1	2.2e+05	13.12	17.00	9.75	0.03
Iceland	Country	IFR	2020-06-14	30-39	1	4.7e+04	1.00	1.50	0.70	2.13
Korea	Country	IFR	2020-07-11	30-39	2	7.1e+06	0.04	0.07	0.02	0.64
New Zealand	Country	IFR	2020-07-09	30-39	0	6.2e+05	0.08	0.12	0.04	0.00
Sweden	Country	IFR	2020-06-18	20-49	63	3.9e+06	6.50	7.84	5.16	0.25
Castiglione d'Adda	Small	IFR	NA	15-64	4	3.1e+03	19.10	23.24	14.86	6.86
Italy Report	Country	Seroprevalence	2020-08-01	35-49	334	1.3e+07	2.40	2.80	2.10	1.10
Spain	Country	IFR	2020-07-15	40-44	78	4.0e+06	3.80	4.60	3.00	0.51
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	40-49	290	8.0e+05	32.20	41.00	23.40	1.13

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who
has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE 2 Complete table of inputs used by the meta-analysis model and crude IFR's. *(continued)*

Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
Gangelt	Small	IFR	NA	35-54	0	3.6e+03	14.00	18.00	12.00	0.00
Iceland	Country	IFR	2020-06-14	40-49	0	4.3e+04	1.50	2.00	1.10	0.00
Korea	Country	IFR	2020-07-11	40-49	3	8.2e+06	0.04	0.06	0.02	0.88
New York	USState	IFR	2020-05-23	40-49	1026	2.4e+06	15.30	17.00	13.70	2.85
New Zealand	Country	IFR	2020-07-09	40-49	0	5.9e+05	0.07	0.11	0.04	0.00
Spain	Country	IFR	2020-07-15	45-49	196	3.9e+06	4.10	5.20	3.30	1.21
Indiana	USState	IFR	NA	40-59	148	1.7e+06	3.14	5.00	1.90	2.81
Western Washington	USState	IFR	NA	40-59	69	1.1e+06	0.90	1.90	0.30	6.88
Spain	Country	IFR	2020-07-15	50-54	230	3.6e+06	4.20	5.20	3.30	1.51
Louisia.	USState	IFR	2020-05-06	50-59	126	5.9e+05	8.30	11.90	4.50	2.59
Minneapolis	USState	IFR	2020-06-04	50-59	47	5.2e+05	0.70	2.80	0.00	13.03
Missouri	USState	IFR	2020-05-23	50-59	43	8.0e+05	2.00	3.80	0.50	2.70
Philadelphia	USState	IFR	2020-05-17	50-64	290	1.1e+06	0.80	2.80	0.00	33.95
Utah	City	IFR	2020-06-19	45-64	24	6.3e+05	2.90	5.20	0.90	1.31
Belgium	Country	IFR	2020-05-16	45-64	280	3.1e+06	6.20	8.30	4.70	1.47
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	50-59	533	6.0e+05	39.10	46.10	32.10	2.27
Connecticut	USState	IFR	2020-05-28	50-59	157	5.2e+05	8.10	11.60	4.80	3.73
Diamond Princess	Small	IFR	2020-04-01	50-59	0	4.0e+02	14.82	14.90	14.70	0.00
Iceland	Country	IFR	2020-06-14	50-59	0	4.2e+04	0.80	1.30	0.50	0.00
Italy Report	Country	Seroprevalence	2020-08-01	50-59	1196	9.6e+06	3.10	3.50	2.70	4.00
Korea	Country	IFR	2020-07-11	50-59	15	8.5e+06	0.06	0.08	0.03	3.20
New York	USState	IFR	2020-05-23	50-59	2764	2.6e+06	16.00	17.50	14.60	6.58

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE 2 Complete table of inputs used by the meta-analysis model and crude IFR's. *(continued)*

Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
New Zealand	Country	IFR	2020-07-09	50-59	0	6.3e+05	0.08	0.12	0.04	0.00
San Francisco Bay	USState	IFR	2020-05-25	50-64	66	1.5e+06	0.70	2.40	0.00	6.37
South Florida	USState	IFR	NA	50-64	169	1.3e+06	2.00	4.00	0.30	6.64
Geneva	City	IFR	2020-06-01	50-64	16	9.9e+04	10.45	14.11	7.31	1.55
Spain	Country	IFR	2020-07-15	55-59	758	3.4e+06	3.90	4.90	3.10	5.69
Sweden	Country	IFR	2020-06-18	50-69	504	2.4e+06	4.81	5.98	3.64	4.38
Spain	Country	IFR	2020-07-15	60-64	1249	2.9e+06	3.50	4.60	2.70	12.14
Gangelt	Small	IFR	NA	55-74	0	3.1e+03	17.00	23.00	12.00	0.00
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	60-69	1155	3.9e+05	40.30	51.40	29.10	7.28
Diamond Princess	Small	IFR	2020-04-01	60-69	1	9.2e+02	19.18	19.30	19.10	5.65
Iceland	Country	IFR	2020-06-14	60-69	2	3.8e+04	0.50	1.00	0.30	10.66
Italy Report	Country	Seroprevalence	2020-08-01	60-69	3274	7.2e+06	2.60	2.90	2.10	17.40
Korea	Country	IFR	2020-07-11	60-69	41	6.5e+06	0.05	0.08	0.03	11.91
New Zealand	Country	IFR	2020-07-09	60-69	3	5.2e+05	0.07	0.10	0.04	8.33
Spain	Country	IFR	2020-07-15	65-69	1905	2.4e+06	4.10	5.30	3.10	19.35
Castiglione d'Adda	Small	IFR	NA	65-74	17	5.4e+02	31.30	37.30	25.40	100.95
Belgium	Country	IFR	2020-05-16	65-74	663	1.1e+06	4.10	7.20	2.30	14.10
Connecticut	USState	IFR	2020-05-28	60+	3633	8.8e+05	4.20	6.00	2.30	98.77
Indiana	USState	IFR	NA	60+	1864	1.5e+06	1.65	2.40	1.00	74.71
Louisia.	USState	IFR	2020-05-06	60+	1053	1.0e+06	4.40	8.00	1.50	23.07
Minneapolis	USState	IFR	2020-06-04	60+	928	8.1e+05	1.00	3.20	0.00	114.11
Missouri	USState	IFR	2020-05-23	60+	620	1.5e+06	3.20	4.60	1.90	13.21

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who
has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE 2 Complete table of inputs used by the meta-analysis model and crude IFR's. *(continued)*

Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
Spain	Country	IFR	2020-07-15	70-74	3230	2.2e+06	3.80	5.10	2.80	38.56
Western Washington	USState	IFR	NA	60+	700	8.7e+05	1.70	2.70	0.90	47.31
New York	USState	IFR	2020-05-23	60+	24376	4.5e+06	12.10	13.10	11.20	44.33
Geneva	City	IFR	2020-06-01	65+	257	8.4e+04	6.82	10.53	3.83	45.09
New Zealand	Country	IFR	2020-07-09	70-79	7	3.6e+05	0.04	0.07	0.02	44.30
Diamond Princess	Small	IFR	2020-04-01	70-79	8	1.0e+03	23.05	23.20	23.00	34.90
Iceland	Country	IFR	2020-06-14	70-79	3	2.3e+04	0.30	1.30	0.27	42.71
Korea	Country	IFR	2020-07-11	70-79	84	3.6e+06	0.05	0.07	0.03	47.89
Philadelphia	USState	IFR	2020-05-17	65+	2374	6.8e+05	1.60	3.50	0.30	219.23
San Francisco Bay	USState	IFR	2020-05-25	65+	333	1.2e+06	0.90	2.50	0.20	30.20
Utah	City	IFR	2020-06-19	65+	71	3.7e+05	2.70	5.00	0.90	7.19
South Florida	USState	IFR	NA	65+	1060	1.2e+06	3.00	4.50	1.70	29.38
Spain	Country	IFR	2020-07-15	75-79	6175	1.8e+06	3.40	5.00	2.40	100.28
Sweden	Country	IFR	2020-06-18	70-95	4485	1.5e+06	3.12	4.12	2.13	94.03
Castiglione d'Adda	Small	IFR	NA	75-84	25	4.0e+02	36.60	44.90	28.30	170.34
Belgium	Country	IFR	2020-05-16	75-84	1182	6.9e+05	7.00	11.70	4.20	24.45
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	70+	2788	3.4e+05	34.30	42.90	25.70	24.07
Gangelt	Small	IFR	NA	75+	9	1.2e+03	12.00	27.00	6.00	63.83
Italy Report	Country	Seroprevalence	2020-08-01	70+	21271	1.0e+07	2.50	2.90	2.10	83.00
Spain	Country	IFR	2020-07-15	80-84	5192	1.3e+06	3.90	6.10	2.40	103.33
Diamond Princess	Small	IFR	2020-04-01	80+	5	2.2e+02	25.00	25.10	24.90	86.42
Iceland	Country	IFR	2020-06-14	80+	4	1.3e+04	0.20	2.50	0.10	156.56

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

TABLE 2 Complete table of inputs used by the meta-analysis model and crude IFR's. *(continued)*

Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
Korea	Country	IFR	2020-07-11	80+	144	1.9e+06	0.06	0.09	0.03	127.21
New Zealand	Country	IFR	2020-07-09	80+	12	1.9e+05	0.04	0.06	0.02	153.85
Spain	Country	IFR	2020-07-15	85+	21248	1.6e+06	2.85	5.60	1.56	455.37
Castiglione d'Adda	Small	IFR	NA	85+	16	1.5e+02	42.10	53.10	31.10	255.07
Belgium	Country	IFR	2020-05-16	85+	1878	3.3e+05	13.20	19.60	8.90	43.55

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

References

[1] 1Day Sooner. 1Day Sooner web site, October 2020.

[2] Gaurav Agrawal, Michael Conway, Jennifer Heller, Adam Sabow, and Gila Tolub. On pins and needles: Will COVID-19 vaccines save the world? Technical report, McKinsey & Company, July 2020.

[3] N Ahmad, N Peterson, and F Torella. The micromort: a unit for comparing and communicating risk to patients. *International Journal of Clinical Practice*, 69(5):515–517, 2015.

[4] Donald A Berry, Scott Berry, Peter Hale, Leah Isakov, Andrew W Lo, Kien Wei Siah, and Chi Heem Wong. A cost/benefit analysis of clinical trial designs for COVID-19 vaccine candidates. Technical report, National Bureau of Economic Research, 2020.

[5] Jennifer Blumenthal-Barby and Peter Ubel. Payment of COVID-19 challenge trials: underpayment is a bigger worry than overpayment. *Journal of Medical Ethics*, 2020.

[6] N. Brazeau, R. Verity, S. Jenks, H. Fu, C. Whittaker, P. Winskill, I. Dorigatti, P. Walker, S. Riley, R. P. Schnekenberg, H. Heltgebaum, T. Mellan, S. Mishra, H. Unwin, O. Watson, Z. Cucunuba Perez, M. Baguelin, L. Whittles, S. Bhatt, A. Ghani, N. Ferguson, and L. Okell. Report 34: COVID-19 infection fatality ratio: estimates from seroprevalence. Report, October 2020.

[7] Angelo Carfì, Roberto Bernabei, Francesco Landi, et al. Persistent symptoms in patients after acute COVID-19. *Jama*, 324(6):603–605, 2020.

[8] Bob Carpenter. Hierarchical Partial Pooling for Repeated Binary Trials, 2016.

[9] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[10] Andrew P Catchpole, Daniel J Fullen, Nicolas Noulin, Alex Mann, Anthony S Gilbert, and Rob Lambkin-Williams. The manufacturing of human viral challenge agents for use in clinical studies to accelerate the drug development process. *BMC research notes*, 11(1):1–6, 2018.

[11] Jonathan J. Deeks. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine*, 21(11):1575–1600, June 2002.

[12] G Miller Franklin and Christine Grady. The ethical challenge of infection-inducing challenge experiments. *Clinical Infectious Diseases*, 33(7):1028–1033, 2001.

[13] Gloria P Gómez-Pérez, Almudena Legarda, Jose Muñoz, B Kim Lee Sim, María Rosa Ballester, Carlota Dobaño, Gemma Moncunill, Joseph J Campo, Pau Cisteró, Alfons Jimenez, et al. Controlled human malaria infection by intramuscular and direct venous inoculation of cryopreserved plasmodium falciparum sporozoites in malaria-naive volunteers: effect of injection volume and dose on infectivity rates. *Malaria journal*, 14(1):306, 2015.

[14] Olivia Grimwade, Julian Savulescu, Alberto Giubilini, Justin Oakley, Joshua Osowicki, Andrew J Pollard, and Anne-Marie Nussberger. Payment in challenge studies: ethics, attitudes and a new payment for risk model. *Journal of medical ethics*, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

[15] Alison Han, Lindsay M Czajkowski, Amanda Donaldson, Holly Ann Baus, Susan M Reed, Rani S Athota, Tyler Bristol, Luz Angela Rosas, Adriana Cervantes-Medina, Jeffery K Taubenberger, et al. A dose-finding study of a wild-type influenza a (h3n2) virus in a healthy volunteer human challenge model. *Clinical Infectious Diseases*, 69(12):2082–2090, 2019.

[16] Ronald A Howard. On making life and death decisions. In *Societal Risk Assessment*, pages 89–113. Springer, 1980.

[17] Euzebiusz Jamrozik and Michael J Selgelid. COVID-19 human challenge studies: ethical issues. *The Lancet Infectious Diseases*, 2020.

[18] Euzebiusz Jamrozik and Michael J Selgelid. *Human challenge studies in endemic settings: ethical and regulatory issues*. Springer Nature, 2020.

[19] Euzebiusz Jamrozik and Michael J. Selgelid. *History of Human Challenge Studies*, pages 9–23. Springer International Publishing, Cham, 2021.

[20] Christian P Larsen, Stephen S Whitehead, and Anna P Durbin. Dengue human infection models to advance dengue vaccine development. *Vaccine*, 33(50):7075–7082, 2015.

[21] Andrew T Levin, Kensington B Cochran, and Seamus P Walsh. Assessing the age specificity of infection fatality rates for COVID-19: Meta-analysis & public policy implications. Technical report, National Bureau of Economic Research, 2020.

[22] Ted Lockhart. *Moral uncertainty and its consequences*. Oxford University Press, 2000.

[23] William MacAskill, Krister Bykvist, and Toby Ord. *Moral Uncertainty*. Oxford University Press, 2020.

[24] David Manheim. Evolving ethics of covid-19 challenge trials. *The Lancet Infectious Diseases*, 2020.

[25] Isaac Martinez, David Manheim, Sean McPartlin, Lena Jewler, Keren Pedersen, Abie Rohrig, Josh Morrison, and 1DaySooner Research Team. Incorporating participant preferences and accounting for power imbalances in human challenge trial ethics (tentative title). *Journal of Practical Ethics*, In Preparation, 2020.

[26] Matthew J Memoli, Lindsay Czajkowski, Susan Reed, Rani Athota, Tyler Bristol, Kathleen Proudfoot, Sarah Fargis, Matthew Stein, Rebecca L Dunfee, Pamela A Shaw, et al. Validation of the wild-type influenza a human challenge model h1n1pdmist: an a (h1n1) pdm09 dose-finding investigational new drug study. *Clinical Infectious Diseases*, 60(5):693–702, 2015.

[27] Harvey Moldofsky and John Patcai. Chronic widespread musculoskeletal pain, fatigue, depression and disordered sleep in chronic post-sars syndrome; a case-controlled study. *BMC neurology*, 11(1):37, 2011.

[28] Millett Granger Morgan, Max Henrion, and Mitchell Small. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge university press, 1990.

[29] Linh Chi Nguyen, Christopher W Bakerlee, T Greg McKelvey, Sophie M Rose, Alexander J Norman, Nicholas Joseph, David Manheim, Michael R McLaren, Steven Jiang, Conor F Barnes, Megan Kinniment, Derek Foster, Thomas C Darton, and Josh Morrison. Evaluating Use Cases for Human Challenge Trials in Accelerating SARS-CoV-2 Vaccine Development. *Clinical Infectious Diseases*, jul 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

[30] CM Nielsen, J Vekemans, M Lievens, KE Kester, JA Regules, and CF Ockenhouse. Rts, s malaria vaccine efficacy and immunogenicity during plasmodium falciparum challenge is associated with hla genotype. *Vaccine*, 36(12):1637–1642, 2018.

[31] World Health Organization et al. Key criteria for the ethical acceptability of COVID-19 human challenge studies. Technical report, World Health Organization, 2020.

[32] Megan O'Driscoll, Gabriel Ribeiro Dos Santos, Lin Wang, Derek A. T. Cummings, Andrew S. Azman, Juliette Paireau, Arnaud Fontanet, Simon Cauchemez, and Henrik Salje. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature*, pages 1–9, November 2020.

[33] Amitabha Palmer and Christopher Schuck. Establishing informed consent for sars-cov-2 challenge trials. Technical report, 1Day Sooner Research Team, 2020.

[34] Ray Perrin, Lisa Riste, Mark Hann, Andreas Walther, Annice Mukherjee, and Adrian Heald. Into the looking glass: Post-viral syndrome post COVID-19. *Medical hypotheses*, 144:110055, 2020.

[35] Meta Roestenberg, Marie-Astrid Hoogerwerf, Daniela M Ferreira, Benjamin Mordmüller, and Maria Yazdanbakhsh. Experimental infection of human volunteers. *The Lancet Infectious Diseases*, 18(10):e312–e322, 2018.

[36] Henrik Salje, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Jehanne Richet, Claire-Lise Dubost, Yann Le Strat, Justin Lessler, Daniel Levy-Bruhl, Arnaud Fontanet, Lulla Opatowski, Pierre-Yves Boelle, and Simon Cauchemez. Estimating the burden of sars-cov-2 in france. *Science*, 369(6500):208–211, 2020.

[37] SK Shah, J Kimmelman, A Drapkin Lyerly, H Fernandez Lynch, F McCutchan, FG Miller, R Palacios, C Pardo-Villamizar, and C Zorrilla. Ethical considerations for zika virus human challenge trials. *National Institute of Allergy and Infectious Diseases*, 2017.

[38] Merrick Pierson Smela, Gavriel Kleinwaks, Nicole Sexton, and Ginny Schmit. Faq: Long-term effects of covid-19, November 2020.

[39] 1Day Sooner, Scott Aaronson, Amrita Ahuja, Chris Anderson, Alexander Berger, Arthur Caplan, Nir Eyal, Ambassador James Glassman, Kim Krawiec, Marc Lipsitch, Abigail Marsh, Josh Morrison, Stanley Plotkin, Sophie Rose, Nadine Rouphael, Sally Satel, Peter Singer, Kate Wharton, and Daniel Wikler. Open letter: Challenge trials for COVID-19, July 2020.

[40] Mark W Tenforde, Sara S Kim, Christopher J Lindsell, Erica Billig Rose, Nathan I Shapiro, D Clark Files, Kevin W Gibbs, Heidi L Erickson, Jay S Steingrub, Howard A Smithline, et al. Symptom duration and risk factors for delayed return to usual health among outpatients with COVID-19 in a multistate health care systems network—united states, march–june 2020. *Morbidity and Mortality Weekly Report*, 69(30):993, 2020.

[41] John J Treanor, Karen Kotloff, Robert F Betts, Robert Belshe, Fran Newman, Dominick Iacuzio, Janet Wittes, and Martin Bryant. Evaluation of trivalent, live, cold-adapted (caiv-t) and inactivated (tiv) influenza vaccines in prevention of virus infection and illness following challenge of adults with wild-type influenza a (h1n1), a (h3n2), and b viruses. *Vaccine*, 18(9-10):899–906, 1999.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.19.20234658; this version posted November 26, 2020.
The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

[42] Robert Verity, Lucy C. Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick G. T. Walker, Han Fu, Amy Dighe, Jamie T. Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma Cucunubá, Rich FitzJohn, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Daniel Laydon, Gemma Nedjati-Gilani, Steven Riley, Sabine van Elsland, Erik Volz, Haowei Wang, Yuanrong Wang, Xiaoyue Xi, Christl A. Donnelly, Azra C. Ghani, and Neil M. Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases*, 0(0), March 2020. Publisher: Elsevier.

[43] Elizabeth J. Williamson, Alex J. Walker, Krishnan Bhaskaran, Seb Bacon, Chris Bates, Caroline E. Morton, Helen J. Curtis, Amir Mehrkar, David Evans, Peter Inglesby, Jonathan Cockburn, Helen I. McDonald, Brian MacKenna, Laurie Tomlinson, Ian J. Douglas, Christopher T. Rentsch, Rohini Mathur, Angel Y. S. Wong, Richard Grieve, David Harrison, Harriet Forbes, Anna Schultze, Richard Croker, John Parry, Frank Hester, Sam Harper, Rafael Perera, Stephen J. W. Evans, Liam Smeeth, and Ben Goldacre. Factors associated with COVID-19-related death using OpenSAFELY. *Nature*, 584(7821):430–436, August 2020.

[44] Witold Więcek and Rachael Meager. baggr: Bayesian Aggregate Treatment Effects package, May 2020.