Adaptations and innovations for vaccine trials in an epidemic: design elements that add value

July 14, 2020

Abstract

A clinical to trial to test a vaccine in the midst of an epidemic faces a number of unique challenges, including low and sparse incidence, difficulty in establishing eligibility of potential participants, and a time pressure to find an efficacious agent in order to curtail an outbreak. Despite these specific issues, vaccine trials in outbreaks often use the same designs as treatment trials. Recently, novel vaccine trial designs were proposed that addressed specific challenges, including targeted recruitment of people at imminent risk of infection, and determination of the trial's size based on the number of events seen rather than the number of people enrolled.

We develop a network-based simulation model to demonstrate and assess these trial designs, focusing on COVID-19 as an example. Restricting our attention to two-arm trials and a binary outcome, we present and assess three design and analysis choices that are geared towards trials for vaccines for infectious diseases that spread person to person, such as COVID-19 and Ebola virus disease:

1. Recruiting contacts who are connected to known cases as a means to maximise number of events seen in a short time horizon.
2. Accounting for the possibility that individuals were infected before a vaccine had a chance to protect them by weighting the data in the analysis (which we call "retrospective exclusion").
3. Determining the size of the trial in terms of the number of events observed, rather than the number of individuals enrolled.

In the treatment-trial literature, adaptive designs are proposed to tailor trials in order to optimise some ethical outcome, such as treatment successes or speed to conclusion. Some such designs are ill suited to vaccine trials due to the different timescales on which they operate. Having identified design choices that allow a vaccine trial to operate on a faster timescale, we overcome that obstacle, and thus enable adaptations that are more familiar in the treatment-trial literature. The adaptations we consider include:

1. Response-adaptive randomisation.
2. Early stopping for efficacy using α spending.
3. Early stopping for futility based on a heuristic to balance expected sample size and power.

We conclude that, together, these design choices allow trials to be designed that meet the specific challenges of vaccine trials, including the ethical objective of maximising population benefit.

Contents

1 Introduction

1.1 Background

Phase 3 clinical trials to test vaccines in the midst of epidemic outbreaks must balance multiple objectives. These objectives include establishing evidence regarding the efficacy of the vaccine, estimating the extent of the protective effect, conferring a health benefit to the trial participants, and conferring a health benefit to the wider population. There are many competing criteria to choose between in designing a trial, and it has been discussed before how choices can be made to prioritise particular objectives (Bellan et al., 2017; Kahn et al., 2018; Nason, 2016; Kahn et al., 2020). For example, two trial elements that can be traded off against each other are the power to detect an effect and the speed with which a conclusion is reached. Metrics have been derived to quantify how ethical designs are, allowing informed comparisons to be made (Bellan et al., 2017). Simulation studies are designed to compare some aspects of trial design, for example whether to randomise at the individual or cluster level (Hitchings et al., 2017), in order to provide evidence for best practice in vaccine trial design. When outbreaks occur, novel designs are proposed and implemented.

Trials for vaccines are currently in progress for SARS-CoV-2 (WHO R&D Blueprint, 2020), (`http://www.ox.ac.uk/news/2020-05-22-oxford-covid-19-vaccine-begin-phase-iiiii-human-trials`), and many vaccines are in development (`www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines`). So far the trial designs proposed have been two-arm, individually randomised placebo-controlled trials, with the exception of the WHO Blueprint, which allows for additional arms to be added. In addition, it plans to carry out statistical tests for efficacy following the observation of a fixed number of cases, rather than following the enrolment of a fixed number of participants (WHO R&D Blueprint, 2020). Participants are (to be) recruited at random from within particular countries in response to incidence (WHO R&D Blueprint (2020), `https://www.ovg.ox.ac.uk/news/trial-of-oxford-covid-19-vaccine-starts-in-brazil`, `https://www.ovg.ox.ac.uk/news/trial-of-oxford-covid-19-vaccine-in-south-africa-begins`).

In the outbreak of Ebola from 2014 to 2016, vaccines were trialled in three studies (Ebola ça Suffit Ring Vaccination Trial Consortium, 2015; Kennedy et al., 2016; Widdowson et al., 2016), which have been reviewed in e.g. Bellan et al. (2017); Kahn et al. (2018); Nason (2016). One trial recruited participants at random (Kennedy et al., 2017), one recruited front-line health workers (Samai et al., 2018), and one recruited according to infection events as they arose (Henao-Restrepo et al., 2017). This latter trial was the only one to conclude efficacy, as too few events were seen in the others (Kennedy et al., 2017; Samai et al., 2018). Its ring recruitment strategy was based on contact tracing, and in this way identified those at imminent risk of infection. In the COVID-19 pandemic of 2020, many states have employed contact tracing as part of their efforts to contain the spread of disease in their countries. Both examples of contact tracing followed international protocols (WHO Disease Surveillance and Response Programme Area Disease Prevention and Control Cluster, 2014; ECDC, 2020).

Elsewhere in the clinical trials literature, and rarely seen in vaccine trials, are adaptive designs. An adaptive design is one in which changes to the trial design occur in a prespecified way as the trial progresses according to data that are accrued, with the aim of improving power, efficiency or participant benefit of clinical trials (Wason et al., 2019). Many elements of a trial might be changed in the course of a trial. The trial described in WHO R&D Blueprint (2020) allows for many planned adaptations, including what the candidate vaccines are, what the control arm is, and who is eligible for recruitment. One of the Ebola vaccine trials included an element of adaptation. The Henao-Restrepo et al. (2017) trial planned for and conducted an interim analysis, and concluded the trial early, having found somewhat convincing evidence of the vaccine's efficacy. Brueckner et al. (2018), on the other hand, have considered the application of adaptations to trials for treatments of Ebola virus disease (EVD).

Among the suite of adaptive designs are interim analyses for early stopping and response-adaptive randomisation (Kahn et al., 2020). An interim analysis is one which is planned for, in which the data accumulated so far are analysed in order to make pre-planned changes to the protocol. "Early stopping" is the decision to terminate the trial concluding efficacy or futility, based on a statistical test or another criterion. When a vaccine is urgent, stopping for efficacy allows the agent to be rolled out more quickly. If resources or potential participants are scarce, stopping for futility releases those resources or potential participants for redeployment in other trials. Adaptation to the randomisation ratio, in which the allocation ratio between experimental and control is adjusted, can be used to maximise power or patient benefit. Either one of these might be a strong motivating force in an outbreak: increasing the number of people vaccinated with a potentially effective vaccine is desirable from the perspective of participant benefit and public health. On the other hand, a high power permits a trial to conclude faster, with fewer patients, which would allow a vaccine to be rolled out more quickly.

To our knowledge Ebola ça Suffit Ring Vaccination Trial Consortium (2015) and WHO R&D Blueprint (2020) are the only examples of vaccine trial designs that include an adaptive feature. We know of no examples of vaccine trial designs including response-adaptive randomisation. Response-adaptive designs require responses that are quick to observe (Wason et al., 2019); such designs have been proposed from treatment trials for infectious diseases, whose responses are often quickly observable, as survival beyond a fixed (small) number of days, or a

change in viral count (Brueckner et al., 2018). For a vaccine, there is no particular time after vaccination where an observable is definitively informative of its success, because success here is the absence of a fail and there is no expected time to failure.

1.2 Open challenges

It has been suggested that adaptive designs might improve a vaccine trial, in terms of efficiency or ethics, through having multiple stages, multiple candidate interventions, and/or adaptive randomisation (Kahn et al., 2020). The choice to employ an adaptive design might have consequences for the other elements of the trial design. First, a design including response-adaptive randomisation requires the endpoint, if binary, (or the event, if time to event) to be near in time. One way to achieve this is to ensure that individuals recruited are at imminent, and therefore high, risk of exposure to infection. The ring vaccination trial (Henao-Restrepo et al., 2017) used such a quickly observable outcome. The follow-up time was 21 days, at which point the participants in the control arm were vaccinated.

Second, designs including response-adaptive randomisation confer largest advantages when testing multiple experimental arms. Design options include the removal of arms for futility, early ending of the trial due to compelling evidence of efficacy, and optimal allocation of individuals to arms with reference to trial and/or patient outcomes (Villar et al., 2015b). Therefore, with an adaptive design, one might favour a multi-arm, multi-stage trial, as proposed in (Kahn et al., 2020).

Adaptive designs are not useful "when increased practical complexity eliminates theoretical efficiency gains" (Wason et al., 2019). A particular challenge for designs including response-adaptive randomisation is accounting for time trends (Villar et al., 2018), in that the design needs to be robust to confounding effects from potential time trends in the data. An important question then is whether the adaptive design is beneficial given the logistical constraints in the setting and the required volume of information. Put another way, can other choices in the design space be made to increase power enough to meet the additional costs associated with adaptive designs? And is it worth it?

1.3 Addressing the challenges

Here, we use a network-based simulation to test, first, how non-adaptive design and analysis choices influence power and efficiency, and, second, whether a superior trial design can be found through adaptation. While response-adaptive randomisation in particular is seen to be most beneficial for multi-arm trials, we explore what differences it makes to two-arm trials, which will indicate what is the "lower bound" for multi-arm trials. Recognising that a quickly observable outcome is crucial to response-adaptive designs (Wason et al., 2019), we focus on recruitment strategies that identify and enrol people at imminent risk of infection.

We explore three trial design and analysis elements: contact tracing vs. random recruitment, as in Henao-Restrepo et al. (2017); the use of time information for downweighting the outcome where it's unclear whether a participant should have been excluded from the outset; and the use of cumulative infection events seen as the way to determine the duration of the trial, rather than setting the "number of participants" at the outset, as in WHO R&D Blueprint (2020). For two of these elements, we identify choices that can be made that increase power for a fixed number of participants. For the final element we identify the different impacts on power the choice to terminate a trial based on number of infection events has compared to the more standard choice of number of participants. These results are important for vaccine trials, where participants might be sparse and infection events rare, and for adaptive designs that try to deliver an "ethical" outcome, which often comes at the cost of power.

We demonstrate how early stopping for efficacy or futility can be integrated into the trial framework and what gains it offers in terms of efficiency. We compare the merits of different response-adaptive randomisation designs. We compare all our proposed design choices to those previously or traditionally used, highlighting choices that increase efficiency (using the minimal resources necessary to reach a conclusion) and information (achieving a higher power for the same trial size, or the same power for a smaller number of participants), and decrease uncertainty in the trial's power, which is crucial for an epidemic whose forecasts come with high uncertainty.

We focus on COVID-19, and make reference also to EVD in illustrating the designs. There are two elements of these diseases that we exploit in our trial design: the way the infection spreads, from person to person; and the scale of the national and international organised responses, in the case of COVID-19 due to the rate at which it spreads, and in the case of EVD due to the high case-fatality rate. This response creates an enabling infrastructure for the trial in terms of resources, protocols, surveillance, contact tracing, personnel, and testing, allowing the trial to achieve more given fixed resources. We anticipate that our results could be generalised to diseases sharing these attributes.

1.4 Outline

In Section 2 we describe our non-adaptive trial design and analysis choices. To this basis, adaptive elements are added in Section 3. We assess and evaluate the designs through simulation, with operating characteristics and summary statistics presented in Section 4, and in Section 5 we conclude. The simulation model and trial mechanics are fully described in Appendix A, and code is available at `github.com/robj411/ADAGIO/COVID19`.

2 Trial design: non-adaptive elements and analysis choices

We use a network model to simulate an epidemic occurring in a population, in which the vaccine trial operates. The model consists of a population from which the trial participants are selected, and who are connected through a contact network, as in Hitchings et al. (2018). It is through these connections that the disease spreads, and each infected individual transitions through the stages of the disease (see Appendix A for details).

The trial design we propose is specifically for an infectious disease with the dynamics and means of spread of EVD or COVID-19; that is, spreading person to person, where it is possible to trace infection events, e.g. through contact tracing, in order to define recruitment criteria for designs that prioritise those at higher immediate risk of infection. In addition, our trial is for a vaccine whose time to seroconversion is fast, that is, a (hypothetical) disease-specific antibody test would go from negative to positive within 15 days, for all participants with high confidence (such as Marzi et al. (2015), `https://investors.modernatx.com/news-releases/news-release-details/moderna-announces-positive-interim-phase-1-data-its-mrna-vaccine`). Such a specification is necessary, as we use seroconversion as a proxy for protection (a vaccine-induced protective immune response) in our simulations. Together, this defines a time constraint for our design: the time to trace contacts and the seroconversion time together must have duration of order no more than that of the spread of the disease. Otherwise, the epidemic outruns contact tracing and therefore the trial designs we present.

We define multiple trial designs through the following choices, where all designs are individually randomised, compare to a placebo control, and use a binary outcome (which is whether or not a person has a PCR-confirmed diagnosis of COVID-19) at a follow-up time of 25 days (see Section 2.3):

1. Retrospective exclusion of participants implemented at analysis points – see Section 2.1
    a. no exclusion
    b. early cases are excluded that may have been infected before randomisation, and therefore should not be protected by a vaccine
    c. downweighting cases by the probability that they were infected before randomisation
    d. downweighting cases by the probability that they were infected before randomisation and estimating the vaccine efficacy simultaneously
2. Recruitment schedule – see Section 2.2
    a. ring recruitment – people who are contacts or contacts of contacts of identified cases, identified through contact tracing, are eligible for recruitment
    b. random recruitment
3. Trial size – see Section 2.4
    a. choosing the trial size based on number enrolled
    b. choosing the trial size based on number of confirmed cases
4. Allocation probability – see Section 3.1
    a. fixed and equal probability of being randomised to the experimental or control arm
    b. adaptively changed during the trial according to a specific response-adaptive rule

2.1 Exclusion criterion implemented at analysis points

Our ideal endpoint would be whether or not a participant became infected after randomisation and vaccine-induced seroconversion, where participants who became infected before are excluded from the trial. For diseases such as EVD and COVID-19, we cannot know the day on which a person is infected, and we cannot know the day on which a vaccinated person seroconverts. However, we can observe whether or not a person is symptomatic, at which point infection can be confirmed through PCR.[1] Therefore, we supplement the primary endpoint – the observed infection status at day 25 – with a retrospective exclusion criterion: we use our knowledge of the disease progression and the day a person becomes symptomatic to weight their inclusion in the analysis.

Infected people remain asymptomatic for some amount of time (the incubation period) before they become infectious and (potentially) symptomatic. The day an enrolled person becomes symptomatic relates to how likely it is that they became infected after randomisation and after vaccine-induced seroconversion through the number

[1] We assume that through surveillance and self-reporting, symptomatic participants correspond to confirmed cases, as was the case in Henao-Restrepo et al. (2017). This means that asymptomatic participants are not counted in our simulation. See Appendix A for more details.

of days elapsed, the estimated incubation period, and the estimated time to vaccine-induced seroconversion. If the incubation time is the estimated time to fever, for example, then the participant (or a person close to the participant) should be asked on which day the fever started. In addition to these factors considered in Henao-Restrepo et al. (2015), one could include also a "reporting time" if it is expected that some time elapses between symptom onset and the reporting of symptoms.

In summary, as we cannot know when someone was infected, we define our primary endpoint as whether or not a person is symptomatic and, in addition, we use the day of symptom onset relative to randomisation to inform a retrospective exclusion criterion. This exclusion of participants manifests as a "weight" between 0 and 1, which we compute following some rules. This challenge is specific to vaccine trials with recruitment of people at imminent risk; the primary measure is usually observed with confidence when participants enter into a treatment trial, or when participants are recruited at random for a vaccine trial. The exclusion criterion is enabled by daily information of PCR-confirmed cases, as in Henao-Restrepo et al. (2017).[2]

We evaluate four choices we might make for the exclusion criterion. In all our evaluations this criterion accompanies the primary endpoint in all aspects of the trial and its analysis: that is, in calculating the effective number of confirmed cases, in calculating response-adaptive randomisation ratios (Section 3.1), and in the final analysis. See Section 4.1.1 for a comparison of the operating characteristics of the different methods, and Appendix B for their mathematical presentation.

2.1.1 No participants are excluded

The most basic analysis would be to apply no retrospective exclusion at all: that is, to include all participants in the final analysis who had PCR-confirmed symptoms of COVID-19 in the 25 days following randomisation. Therefore, no additional information is required, and in terms of the weighted analysis every case has a weight of 1.

When all cases are included, we are effectively testing "what is the effect of the intervention", which is less than the effect of the vaccine, as some people vaccinated cannot benefit from the vaccine. It might be interesting to know in its own right, for example to test what the effect might be of vaccinating in addition to contact tracing as a method for containment of the outbreak (Fyles et al., 2020).

2.1.2 Exclusion criterion – binary weighting

We define an exclusion rule analogous to that employed in Henao-Restrepo et al. (2017), which is to say we exclude any participant that displays symptoms within a certain number of days of randomisation.The difference between this method and the analysis in which no participants are excluded is that here the effect of the vaccine is isolated from the effect of the intervention.

In terms of the weighted analysis, participants who show symptoms within nine days of randomisation are excluded absolutely and all others are included absolutely, corresponding to binary weights of zero and one. The additional information required is whether the day of symptom onset was before or after this pre-specified threshold, corresponding to the estimated distribution of the incubation period for COVID-19 and an assumed distribution for the time to seroconversion in response to the vaccine.

2.1.3 Downweighting – continuous weighting

From presumed knowledge of infection dynamics, we can allocate continuous weights to observations, rather than binary weights corresponding to absolute inclusion or exclusion. With this method we define a participant's weight to be the probability that they were infected after randomisation and vaccine-induced seroconversion. As before, participants who are not confirmed cases are included in full.

The probability that a person whose symptoms began after vaccination (and seroconversion) was infected after seroconversion depends on the effect of the vaccine. If the vaccine is effective, then this probability is smaller than it would be if the vaccine had no effect. Therefore, as an additional development, we estimate the vaccine efficacy and the weights together. See Figure 1 and Table 1 for illustrations, and Appendix B for the derivation. The results for this weighting method in terms of power and type 1 error are shown alongside the other exclusion methods in Section 4.1.1.

[2]PCR at the time of symptoms would be used preferentially to antibody testing due to the difficulty in distinguishing between antibodies in a person at the end of trial signalling past infection and those signalling vaccination (Dean et al., 2019), and because seroconversion following infection may occur many days after onset of symptoms (Long et al., 2020).

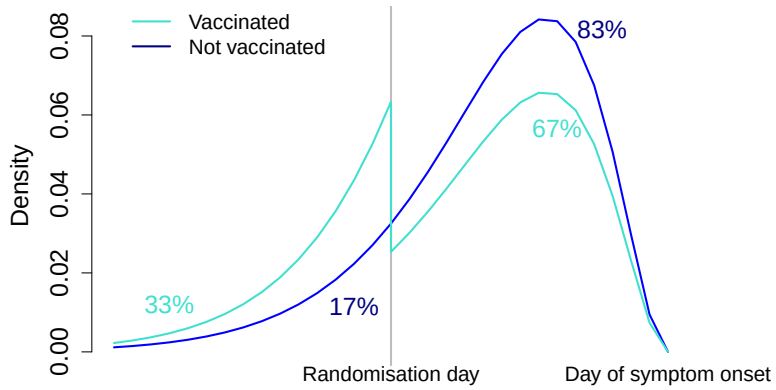

Figure 1: Calculating the inclusion weight given a vaccine effect. Two participants show symptoms on the same day. One (turquoise) is vaccinated. The other (navy) isn't. Given the distribution of the incubation period, the navy person has a probability of 0.83 of having been infected after randomisation day. Their weight is therefore 0.83. Given the vaccine efficacy ($VE > 0$), the turquoise person is less than 83% likely to have been infected after randomisation day.

Table 1: Examples of weights for the different exclusion methods, considering the day of symptom onset (8 or 10 days), whether the participant was vaccinated or control, and the estimated vaccine efficacy (VE, 0 or 0.6).

	Symptoms on day 8				Symptoms on day 10			
	Control		Vaccinated		Control		Vaccinated	
Method	VE=0	VE=0.6	VE=0	VE=0.6	VE=0	VE=0.6	VE=0	VE=0.6
No exclusion	1	1	1	1	1	1	1	1
Exclusion	0	0	0	0	1	1	1	1
Downweight without VE	0.5	0.5	0.5	0.5	0.83	0.83	0.83	0.83
Downweight with VE	0.5	0.5	0.5	0.29	0.83	0.83	0.83	0.66

2.2 Recruitment schedule: contact tracing

We use the recruitment strategy of the ring of contacts employed in Henao-Restrepo et al. (2017), which was inspired by the role of ring vaccination in international elimination of smallpox (Fenner et al., 1988). Participants are eligible for enrolment when someone in their contact network is confirmed as a "case", starting from the first day of the trial.[3] We define the ring as consisting of contacts and contacts of contacts, where a contact-of-contact is to a contact as the contact is to the index case. These contacts are found through "contact tracing", as described in ECDC (2020). In the terminology of Fyles et al. (2020), our rings, consisting of two levels of contacts, come from "two-step contact tracing". In the terminology of graph theory, the "ring" is the neighbourhood of order 2 of the index case. Recruiting participants who are at imminent risk of infection is akin to the method of recruitment of high-risk participants, which was the motivation for Samai et al. (2018) to recruit front-line health workers in the Ebola outbreak of 2014–2016.

We distinguish between different types of relationships (or edges) in the network. We define a "predictable" edge as a relationship that will be discovered through contact tracing, and the "predictable fraction" of a network as $r_s = e_s w_s/(e_s w_s + e_u w_u)$, the weighted sum of predictable edges over the total weighted sum of edges, where there are e_s predictable and e_u unpredictable edges, with weights w_s and w_u respectively.[4] The notion of unpredictable edges here is related to "environmental transmission" defined in Ferretti et al. (2020), that is, through the environment and not from a close contact. The notion of a "predictable" edge here is like the definition of an "acquaintance" in Kucharski et al. (2020) as someone who has been "met before". In our work, the predictable fraction is entirely determined by how "contact tracing" is defined: any contact that is not traced is by definition unpredictable. For our simulation, we specify three types of edge (household, workplace, and random) and stipulate that household and workplace edges are predictable and have weight 1, while random edges are unpredictable and have weight 1/10. (See Appendix A for details of the network.)

Section 4.1.2 illustrates the value of using ring recruitment through contact tracing for recruiting from a group at imminent risk. First, we demonstrate the relationship between the predictable fraction and the trial's power. We adjust the predictable fraction by adjusting the edge weights, rather than, say, the number of edges, aiming to maintain the same overall total sum of edge weights. As the predictable fraction decreases, the disease transmission tends towards spread along random connections rather than through identifiable relationships. In this case recruiting by contact tracing is unlikely to be more helpful than random recruitment. In contrast, where the predictable fraction is high, contact tracing is likely to helpfully anticipate imminent cases, and will likely return a higher-power trial than one that recruits participants randomly. We therefore also compare the ring recruitment method to a random recruitment method where the predictable fraction is 0.91, which is the value from our standard network (see Appendix A).

2.3 Follow-up time

To compare all the designs side by side, we use the single follow-up time of 25 days, which is the day after randomisation on which we follow up a participant for the final time. This is similar to the follow-up time of 21 days used for EVD (Hitchings et al., 2017; Henao-Restrepo et al., 2015). A natural lower bound for the follow-up time given our use of ring recruitment is two "serial intervals" (which is less than 20 days with credibility 99.9% for COVID-19 (He et al., 2020)). Two serial intervals allow for a first contact to become symptomatic within the duration of one serial interval, and the second to follow after, where we allow for the first contact to have become infected before enrolment but the second, after enrolment. We use these heuristics to estimate a window of time in which we are most confident of observing events. We can use simulation to narrow it down further, based on desired operating characteristics (Hitchings et al., 2017).

Our follow-up time of 25 days is used to define the primary endpoint both for adapting (as described in Section 3.1) and for the final analysis at the end of the trial. We fix the follow-up time for reasons of simplicity and computational efficiency, in that we simulate a network only up to day 25. In application, one could instead use as the endpoint for the final analysis the outcome at the end of the trial, that is, we could use the data that continue to accrue over time, beyond 25 days after each participant was enrolled. Inclusion of data up to the end of the trial would likely improve power for all methods, depending on the circumstances: Hitchings et al. (2017) show that, in a declining epidemic, where the effective reproduction number is below 1, there is a limit to the return on extending the follow-up time for a ring-recruitment vaccine trial.

[3]Note that here we use "ring" with reference to the method for recruitment. In our work this is distinct from the randomisation method, which, in other implementations of the strategy, has been cluster randomisation (Henao-Restrepo et al., 2015).

[4]The edge weight is a property of the network that we use to convey the extent of contact, which translates to the relative propensity to transmit infection through this contact. For example, if we specified edge weights of 1 for all relationships then all would be equally likely to transmit infection.

2.4 How large should the trial be?

We propose an event-driven (Schoenfeld, 1983) (or information-based (Mehta et al., 2009), or endpoint-driven (WHO R&D Blueprint, 2020)) trial design. That is, we propose that the trial is terminated following the accumulation of a prespecified number of confirmed cases, rather than a prespecified number of enrolled participants, corresponding to a fixed sample size. (N.B. as all confirmed cases are weighted, the number of cases is in fact the effective number of cases, and the sample size is the effective sample size.) We compare this to a fixed-sample-size design in terms of the power: do different possible epidemic realisations result in predictable or uncertain powers for the trial designs under the different rules governing the final trial size?

In Section 4.1.3, to assess the utility of defining trial size according to cumulative observed cases, we simulate three scenarios: one, where we the parameters for the simulated trials are the same as those for the trial design, and ones where the rate of transmission for the simulation is 4/5 and 5/4 what was used for the design. Varying the rate of transmission serves to deviate the trajectory of number of participants and cases observed systematically from what is expected. This allows us to show how the number of participants and the number of cases each affect power, which is suggestive of what happens to a trial's power when its number of participants is fixed and when its number of cases is fixed. We additionally compare the two designs where there is no vaccine efficacy, demonstrating how many resources are consumed in order to conclude futility.

Note that this artificial discretisation of the outcome space into three represents the variation we might see due to stochasticity, or due the model systematically under or overestimating the number of cases. Indeed, in a trial in an unfolding epidemic, one would likely not be able to tell whether, or to what extent, the deviance is due to stochasticity and to what extent it is due to the model being systematically wrong. Therefore, we use different model parametrisations to investigate the effects of the different epidemic realisations on power subject to the design of the trial.

3 Adaptive designs: interim analyses

We present two adaptations for vaccine trials: early stopping for efficacy or futility, and response-adaptive randomisation. In changing the allocations of participants to arms as the trial progresses in response to accumulated data, we compare designs that prioritise placing more participants in better-performing arms to designs that prioritise maximising power per participant.

3.1 Adaptive allocation probabilities

In the response-adaptive trial simulations, at fixed-frequency, pre-specified moments in the course of the trial, we update the allocation probabilities, using the formulae presented in Sections 3.1.1 and 3.1.2. The allocation probabilities are the probabilities of individuals being allocated to each arm of the trial. We choose the frequency of adaptations to coincide with the follow-up time – that is, 25 days – and all data accrued up to the adaptation day (up to the follow-up time of 25 days post randomisation) are used in computing the probabilities.

To compute the allocation probabilities, first we define π_v as the probability of being allocated to arm v, where v is 0 for control and 1 for experimental. Below we present two families of methods to compute π_1: frequentist and Bayesian. Both require first generating estimates $\hat{p}_0$ and $\hat{p}_1$ of p_0 and p_1, the probability to be uninfected at the follow-up time for the control and experimental arms, respectively. We compare all methods to a design with fixed and equal allocation probabilities, denoted FR (fixed randomisation), in which $\pi_0 = \pi_1 = 0.5$ for all time.

3.1.1 Frequentist response-adaptive

In the frequentist framework, the probabilities are calculated as the maximum-likelihood estimate (MLE): the probability is the number of successes over the total number of observations; $\hat{p}_v = 1 - f_v/N_v$, where f_v is the weighted number of "fails" for arm v, and N_v the sample size for the arm (Hu and Rosenberger, 2006).

Ros We define $\rho = \sqrt{\frac{\hat{p}_1}{\hat{p}_0}}$ as the allocation ratio of experimental to control, as in Rosenberger et al. (2001). Then the allocation probability to the experimental arm is

$$\pi_1 = \frac{\rho}{\rho + 1} = \frac{\sqrt{\hat{p}_1}/\sqrt{\hat{p}_0}}{\sqrt{\hat{p}_1}/\sqrt{\hat{p}_0} + 1} = \frac{\sqrt{\hat{p}_1}}{\sqrt{\hat{p}_1} + \sqrt{\hat{p}_0}}.$$

For example, if $\hat{p}_0 = 0.2$ and $\hat{p}_1 = 0.8$, then $\rho = \sqrt{4} = 2$, and $\pi_1 = 2/3$. This method is designed to balance power and patient benefit, in terms of maximising treatment successes. In the multi-arm case, worse-performing experimental arms create slack for this optimisation. For the two-arm case, then, we expect to see little overall gain, as there is only one experimental arm.

Ney We use the Neyman frequentist rule, which is designed to improve power compared to equal allocation (Rosenberger et al., 2001), by setting

$$\pi_1 = \frac{\sqrt{\hat{p}_1(1-\hat{p}_1)}}{\sqrt{\hat{p}_0(1-\hat{p}_0)} + \sqrt{\hat{p}_1(1-\hat{p}_1)}}.$$

For $\hat{p}_0 = 0.2$ and $\hat{p}_1 = 0.8$, this rule has $\pi_1 = 0.5$. (For the case $\hat{p}_0(1-\hat{p}_0) = 0$ or $\hat{p}_1(1-\hat{p}_1) = 0$, we set $\pi_0 = \pi_1 = 0.5$.)

3.1.2 Bayesian response-adaptive (Thompson sampling)

At the adaptation point, we obtain the posterior distribution of p_i given a uniform prior and the observed data as $\text{Beta}(1 + N_v - f_v, 1 + f_v)$. The allocation probability π_i is defined in terms of these posterior probabilities, computed by sampling, as

$$\pi_1 = \frac{\Pr(p_1 > p_0)^\phi}{\Pr(p_1 > p_0)^\phi + \Pr(p_1 < p_0)^\phi},$$

where we define a tuning parameter ϕ. For Thompson sampling (TS), we set $\phi = 1$, so π_1 is just $\Pr(p_1 > p_0)$, and for TS with tuning (TST), $\phi = j/e$, with j the day of the current update, and e the trial's expected total duration, e.g. 100 days. ϕ therefore takes the value 0 at the beginning of the trial and goes to 1 as the trial progresses, so that the speed with which the allocation probability can reach extreme values (0 or 1) is tempered.

The Thompson sampling methods have a possibility of generating very high allocation probabilities. While tuning slows this process, we find that the TST method tends to 1 over a few adaptations (results not shown). We therefore bound the allocation probability above at 0.8 and below at 0.2 for all methods. Additionally, we terminate the trial and conclude efficacy if and when we compute an allocation probability of 0.99.

3.2 Response-adaptive randomisation and time trends

The epidemic unfolding in real time gives rise to time trends in incidence of the disease, which in turn will result in patient drift in the trial. By "patient drift", we mean some effect that changes over time and influences all participants regardless of vaccination status (as opposed to "temporal trends", which we define to be a time-dependent effect that will affect some but not all arms in the trial, and which we do not consider here). For instance, a natural increase or decline in incidence, or a step change due to government policy on social contact, as proposed by Ferguson et al. (2020), or even a change in the recruitment strategy, would induce a time trend in the epidemic. As both the adaptive trial design and the epidemic are changing over time, in analysing the results, we account for time dependencies of disease exposure in order to infer the effect of the experimental vaccine.

We use randomisation-based inference to correct for patient drift as described by Simon and Simon (2011), which involves resampling the data in order to generate a new null distribution for the test statistic to which to compare the one we compute. The resulting powers and rates of type 1 error are presented in Section 4. In the Appendix, we demonstrate that the correction preserves type 1 error at the cost of a little power in the context of a strong downward linear trend in incidence.

3.3 Early stopping

In Section 4.3, we present simulation results for stopping early. We stop early for efficacy if we can reject the null hypothesis of no vaccine effect at the interim analysis point given the results accrued so far. We choose one interim analysis point, and follow the α spending strategy (Pocock, 1977; O'Brien and Fleming, 1979; Henao-Restrepo et al., 2017). We split α (the threshold for the test statistic below which we conclude efficacy) between the two analysis points – the interim analysis and the final analysis. We choose 0.03 for the first test, leaving 0.02 for the second. We try three event-driven triggers for the interim analysis: testing when the effective number of cases observed is 12, 14, or 16, with the final analysis then taking place at when the effective number of cases is 30.

We stop early for futility if the number of cases in the experimental arm at the interim analysis is high enough that concluding efficacy at the final analysis would be "unlikely". What is deemed "unlikely" is subjective. The value chosen for this threshold will determine the probability that the trial stops early, and it will influence the power. We use an effective number of cases of 8 at each potential trigger point to demonstrate the effect on the trial's power and expected number of participants.

4 Results

We demonstrate the methods described in Sections 2 and 3 using the model and trial system described in Appendix A. The results represent 1,000 simulated trials (unless otherwise stated), where each "trial" involves independent networks (as many independent networks as required to achieve a particular total number of weighted cases), or a fixed number of independent networks, as stated. The "base case" for our simulations is one where the exclusion criterion uses a continuous weighting accounting for the vaccine efficacy, the allocation probabilities are fixed and equal, the follow-up time is 25 days, the target weight for confirmed cases is 24, and the vaccine efficacy assuming a positive effect is 0.7. We report operating characteristics including the number of people enrolled and the number of those symptomatic (which we equate to a confirmed case), the power, the estimated vaccine efficacy (VE), and the type 1 error rate, alongside the details of the design. The duration is reported in days assuming an average of participants enrolled per day.

The power (the probability to correctly reject the null hypothesis) and type I error rate (the probability to incorrectly reject the null hypothesis) are estimated as the proportion of simulations under the positive effect and the null, respectively, for which the null hypothesis was rejected (see Appendix B for mathematical details). The effect of the vaccine is estimated using all simulations under the positive effect (whether or not the trial realisation concluded efficacy). The other columns of the tables (e.g. the numbers of people enrolled and symptomatic) are computed under the alternative unless stated otherwise.

Additionally, we report the "prevented export infections", defined as the expected number of infection events of non-contact-network people for 100 contact networks with no vaccine effect, minus the equivalent number given a vaccine effect of 0.7. An example interpretation for e.g. 10 prevented export infections would be "if we vaccinate $(70\pi)\%$ of people in 100 contact networks, we expect to avert ten infection events from those populations to the outside population in a 25-day period, compared with not vaccinating anyone" (where π is the allocation rate, and is 0.5 for fixed and equal randomisation, and our simulations assume an enrolment rate of 70%). While we don't expect this to be predictive of actual numbers of infections observed, the relative numbers between methods are indicative of the trial's possible or probable effects on the wider epidemic, which would be a useful feature to develop in order to enable assessment of the impacts of trial designs.

4.1 Non-adaptive design and analysis choices

4.1.1 Exclusion criterion implemented at analysis points

In Table 2, we compare the different methods presented for defining the retrospective exclusion criterion. Where no weighting is used ("None": all confirmed cases are included equally), the trial is short but the power is very low. A big gain in power is made by excluding early cases ("Binary"), along with enrolment of more participants. By downweighting ("Continuous, without VE"), there is a slight further increase in power, and finally, by computing the vaccine efficacy alongside participant weights ("Continuous, with VE"), the power is highest, and the VE estimate is closer to the true value of 0.7. The gain in power is due to the correct accounting for the vaccine efficacy when determining which early cases are likely to have been infected before randomisation. The other operating characteristics are similar between the three weighted methods. Computed numbers of participants for one trial are shown in Table 10 in the Appendix, including the "true" number of vaccinated participants who were infected after seroconversion, and control participants who were infected after recruitment.

4.1.2 Recruitment schedule: contact tracing

In Table 3 we consider five possible values for the "predictable-to-unpredictable contacts" balance, realised by adjusting the edge weights in the network. The results show the extent to which power diminishes as the fraction of predictable edges decreases. When more contacts are predictable, more contacts are recruited into the trial, and the number of symptomatic people (confirmed cases) in the trial increases. This is a result of the recruitment method better targeting those at risk. Notice also that the vaccine better protects people when the fraction is higher, in that the difference between the number symptomatic under the null and under the effective vaccine increases as the fraction increases.

In Table 4, where recruitment is random, rather than through contact tracing, many more participants must be recruited in order to observe the requisite number of events. In addition, many more people in the general population must become symptomatic in order for us to observe the requisite number of cases in our random, non-targeted trial population, which is reflected in the higher number of prevented export infection events using the ring recruitment strategy.

Table 2: Comparison of designs where the binary endpoint has no weight, a binary weight, or a continuous weight that is independent of the estimated vaccine efficacy, or is calculated simultaneously. Participants are recruited following the ring strategy. The trial follows the FR design with a follow-up time of 25 days. The trial ends when an effective number of 24 cases have been observed. Standard deviations in brackets. Note that in our simulations we enrol whole contact networks and follow them up for 25 days, so 24 will be the lower bound, and this might be overshot in the case of no weighting when a single network returns multiple cases. (E.g. suppose we reach a weight of 24 on day 10 of the last enrolled contact network; we arrive at a weight of 27 if in the remaining days up to day 25 we observe a further three cases.)

Weighting	Number of participants	Symptomatic	Vaccinated	Power	Type 1 error	VE estimate	Number of participants (null)	Prevented export infections
None	1069 (326)	25	534	0.39	0.05	0.38 (0.27)	804 (274)	6.18
Binary	1725 (502)	41	862	0.71	0.04	0.59 (0.19)	1091 (372)	5.85
Continuous (without VE)	1844 (519)	44	922	0.73	0.03	0.59 (0.17)	1160 (381)	5.79
Continuous (with VE)	1917 (541)	45	958	0.81	0.04	0.65 (0.18)	1158 (383)	5.41

Table 3: The effect of the predictable fraction of transmission events on power. The trial follows the FR design with a follow-up time of 25 days and ends once 100 contact networks have been enrolled and weights exclusion with a continuous variable that accounts for VE. Participants are recruited following the ring strategy. Standard deviations in brackets.

Predictable fraction	Number of participants	Symptomatic	Vaccinated	Power	Type 1 error	Symptomatic (VE=0)	VE estimate
1.0	2913 (169)	83	1455	0.95	0.05	106	0.65 (0.14)
0.8	2900 (172)	62	1449	0.86	0.06	79	0.65 (0.18)
0.6	2912 (169)	45	1455	0.71	0.06	55	0.62 (0.25)
0.4	2908 (165)	30	1453	0.53	0.07	36	0.59 (0.43)
0.2	2912 (171)	19	1454	0.35	0.07	21	0.23 (4.97)

Table 4: Comparison of designs where participants are recruited following the ring strategy vs. recruited at random. The trial follows the FR design with a follow-up time of 25 days and weights exclusion with a continuous variable that accounts for VE. The trial ends when an effective number of 24 cases have been observed. Standard deviations in brackets.

Recruitment	Number of participants	Symptomatic	Vaccinated	Power	Type 1 error	VE estimate	Prevented export infections
Random	12381 (2418)	29	6191	0.78	0.05	0.65 (0.19)	2.16
Ring	1920 (540)	45	960	0.80	0.04	0.65 (0.18)	4.99

Table 5: Comparison of designs where the trial size is determined by the number of cases to designs where the trial end is determined by the number of people enrolled. We make comparisons under three scenarios: one, where we assume that the parameters we estimated were correct ("True"), and ones where the true rate of transmission is 4/5 ("Lower") and 5/4 ("Higher") what we estimated. Participants are recruited following the ring strategy. The trial follows the FR design with a follow-up time of 25 days and weights exclusion with a continuous variable that accounts for VE. Standard deviations in brackets.

True beta	Trial size determined by...	Number of participants	Duration (days)	Symptomatic	Vaccinated	Power	VE estimate	Prevented export infections	Number of participants (no vaccine effect)	Type 1 error
True	Cases	1920 (540)	85 (17)	52 (8)	960	0.80	0.65 (0.18)	4.99	1192 (341)	0.04
Lower	Cases	3017 (793)	119 (25)	57 (10)	1508	0.79	0.64 (0.2)	2.34	1836 (548)	0.04
Higher	Cases	1235 (332)	63 (10)	48 (7)	617	0.81	0.64 (0.18)	9.88	894 (144)	0.04
True	Number of participants	1939 (115)	85 (4)	52 (12)	970	0.81	0.65 (0.19)	4.94	1939 (115)	0.05
Lower	Number of participants	1939 (115)	85 (4)	37 (9)	970	0.62	0.63 (0.28)	1.97	1939 (115)	0.06
Higher	Number of participants	1939 (115)	85 (4)	76 (16)	970	0.91	0.65 (0.14)	9.68	1939 (115)	0.05

Table 6: Comparison of response-adaptive designs. The outcome has a continuous VE-corrected weighting. Participants are recruited following the ring strategy. The follow-up time is 25 days. The trial ends when 24 effective cases have been observed. Standard deviations in brackets.

Adaptation	Number of participants	Duration (days)	Symptomatic	Vaccinated	Power	Power (corrected for time trend)	Type 1 error	Type 1 error (corrected for time trend)	VE estimate	Prevented export infections
Ney	1768 (481)	80 (15)	50	747	0.82	0.76	0.05	0.04	0.67 (0.19)	3.97
Ros	1923 (543)	85 (17)	52	969	0.82	0.80	0.04	0.03	0.65 (0.19)	5.00
TST	2032 (638)	88 (20)	51	1261	0.77	0.76	0.04	0.04	0.64 (0.19)	6.62
TS	1799 (740)	81 (23)	45	1148	0.80	0.74	0.04	0.05	0.67 (0.21)	5.99
FR	1920 (540)	85 (17)	52	960	0.80		0.04		0.65 (0.18)	4.99

4.1.3 How large should the trial be?

In Table 5, we compare a design where we terminate the trial once we have recruited a fixed number of contact networks (60) to one that terminates once we have observed a minimal number (24) of events. In our "base case", using the parameters described in the Appendix, the two approaches have the same expected power (0.8). The expected number of cases for 60 contact networks is 24, and the expected number of contact networks for 24 cases is 60. Differences emerge when we compare the case where the vaccine has no effect, and where the outcome diverges from the expected values.

Where there is no vaccine effect, the fixed-number-of-participants method enrols ≈800 more participants than the fixed-case method on average. Similarly, if we observe more events per participant than we expected under the experimental, the number of participants for the fixed-number-of-participants method enrols ≈700 more participants than the fixed-case method on average. Conversely, if we see fewer events than we expected, the fixed-number-of-participants method concludes underpowered with 1,939 participants, and the fixed-case method enrols 1,000 more participants but maintains power.

This presents another trade-off, as described in Mehta et al. (2009): if the trial size is defined by a pre-determined number of participants, we observe a range in numbers of cases, and a range in power. If the trial size is defined by a pre-determined number of cases, we observe a range in numbers of participants, but the power remains constant. Put another way, fixing sample size gives rise to uncertainty in the power, and fixing case numbers gives rise to uncertainty in the sample size, but the power can be expected to remain around 0.8, whatever happens in the epidemic.

4.2 Response-adaptive randomisation

The results for a suite of adaptive trials are shown in Table 6. We report the power corrected for the time trend in the epidemic following Simon and Simon (2011). See Figure 5 for operating characteristics under different trends and Figure 4 for trajectories of allocation rates.

The Thompson sampling methods (TST and TS) allocate more participants to vaccination instead of control. As a result, there are fewer infections in these trials, fewer infections exported from the network, and the power is lower. The Neyman method has lower power also because of unequal allocations, in this case with more participants in the control arm. The Rosenberger method is most similar to the fixed and equal randomisation design, in terms of number of participants, participant allocation and power.

The trade-off between power, number of people vaccinated, and the time taken to conclude efficacy is the crux of this design choice. In Table 6, we fixed the number of cases observed in the trial population to a total weight of 24, which keeps power close to 0.8. We can trade off number of people vaccinated against time to conclude directly: the fastest trial vaccinates the fewest people and sees the most export infections; the longest trial vaccinates the most people and sees the fewest export infections.

Another means to compare the methods would be to ask what can the trial achieve within, say, 100 days, assuming an efficacious vaccine, and where the trial lasts at most 85 days. We assume that the trial begins when the first participant is enrolled and ends when last participant has been followed up for 25 days. Assuming one contact network can be enrolled per day, 60 contact networks are enrolled. Table 7 shows this comparison and that, when set up with a limited time, the trials that allocate most participants to the efficacious experimental arm – TS and TST – are most ethical, in terms of vaccinating the most people in a fixed time of 100 days. (There are also fewer export infection events in the course of the trial – both in total and per day.) However, beyond 100 days, the designs with higher powers look better, as they are better able to identify an efficacious vaccine, and therefore will vaccinate more people in the long run.

Together, Tables 6 and 7 suggest that where participants are abundant, a trial can justify, in terms of power, allocating more participants to the best performing arm, which would imply vaccinating more people (if the vaccine seems promising from data so far). Where participants are abundant, the epidemic is spreading, and so vaccinating more people would have the desirable side effect of attenuating the epidemic to some extent. On the other hand, where the trial is limited by the number of participants it can recruit, as it would be in a declining epidemic, a trial that maximises power would be more ethical. By not prioritising vaccination, it looks less ethical in a short time horizon (100 days), but is more ethical at longer time horizons (200 days and beyond), as an efficacious vaccine is more likely to be identified by prioritising information gain.

4.3 Early stopping

4.3.1 By interim analysis

As described in Section 3.3, we plan one interim analyses with threshold $\alpha = 0.03$, and a final analysis with $\alpha = 0.02$ with 0.8 power. We choose 30 as the final effective number of cases at which to test with $\alpha = 0.02$, which gives us a power of 0.8 for a range of total weights. We compare different rules for when the interim analysis is performed, defined by different effective numbers of cases observed, with $\alpha = 0.03$. We compare

Table 7: Comparison of response-adaptive trials that last at most 85 days. We compare their ethical profile in terms of the totals vaccinated up to days 100 and 200, which are sums of "Vaccinated in trial" and vaccinated after the trial up to day 100, and up to day 200. "Vaccinated in trial" is the expected number vaccinated during the trial, and "Vaccinated up to day 100 (200)" is the expected number vaccinated after the trial up to day 100 (200), estimated as 32 people vaccinated per day, for the days remaining after the end of the trial, multiplied by the power (the probability to have concluded efficacy and rolled out the vaccine). Standard deviations in brackets.

Adaptation	Sample size	Duration	Symptomatic	Vaccinated in trial	Power	Export infections in trial	Vaccinated up to day 100	Vaccinated up to day 200
Ney	1939 (115)	85 (4)	55	807	0.76	25	1177	3640
Ros	1939 (115)	85 (4)	52	977	0.79	24	1360	3910
TST	1787 (254)	80 (8)	45	1111	0.72	21	1574	3889
TS	1611 (430)	75 (13)	40	1001	0.66	20	1533	3661
FR	1939 (115)	85 (4)	52	970	0.81	24	1363	3981

the interim-analysis designs with not conducting an interim analysis, ending with 24 effective cases (Table 8, "*Assuming VE=0.7*").

Table 8: Trial operating characteristics for different interim-analysis choices, where we test whether to stop for efficacy or futility when the effective number of cases reach a certain size. The bottom row shows the expectation when there is no interim analysis, the trial ends at 24 effective cases, and is tested with $\alpha = 0.05$. In all cases type 1 error is controlled at 0.05. Duration is measured in days. Standard deviations shown in brackets.

Effective number of cases	Probability of stopping early... for efficacy	for futility	Final power	Expected duration (days)	Expected duration of trial that stops early	Expected duration of trial that doesn't stop early
Assuming VE=0						
12	0.04	0.25	0.06	40 (18)	20 (9)	46 (13)
14	0.04	0.42	0.06	37 (18)	22 (9)	46 (13)
16	0.04	0.60	0.05	34 (16)	25 (10)	47 (13)
	0	0	0.04	37 (12)		
Assuming VE=0.7						
12	0.48	0.01	0.81	56 (28)	32 (13)	74 (18)
14	0.53	0.02	0.80	56 (26)	37 (13)	74 (18)
16	0.58	0.04	0.81	56 (24)	41 (14)	74 (18)
	0	0	0.8	60 (16)		

4.3.2 Stopping for futility

What effective case size of the experimental arm at the interim-analysis exceeds the minimum value for which we might conclude efficacy at the final analysis? For a range of final numbers of participants (1,400 to 2,600), a weight of 10 or higher for the experimental arm is sufficient to disable rejection of the null hypothesis for a test at $\alpha = 0.02$ and 30 total weighted cases (which are our assumed conditions for the final analysis). This is a particularly conservative choice, and would result in stopping early at most 10% of the time for futility if there is no vaccine effect. We demonstrate what a less conservative choice of stopping for a weight of 8 or higher looks like, where there is a high probability to stop for futility where there is no vaccine effect (Table 8, "*Assuming VE=0*"), and a low probability to stop early for futility where there is a vaccine effect (Table 8, "*Assuming VE=0.7*").

5 Discussion

We have considered designs for vaccine trials in epidemic outbreaks, focusing on highly infectious, person-to-person–spread disease, using the example of COVID-19. Because the disease spreads through person–person contacts, it is possible to anticipate who is at high and imminent risk of becoming infected. Therefore we use the ring recruitment strategy for this disease.

Our simulation differs from the ring vaccination trial for EVD as described in Henao-Restrepo et al. (2015) in a number of ways. We simulate an iRCT, whereas their trial was a cRCT. One consequence of this is they had stricter criteria for eligibility: in a cRCT, a person is enrolled only if their cluster is enrolled. In our simulation, each person's enrolment is independent. So, a person identified in a contact network with 60% or more overlap with an enrolled contact network would not be enrolled in a cRCT, but they can be enrolled in an iRCT. However, we expect our methods to be otherwise compatible with a cRCT design.

Some people test positive for SARS-CoV-2 but show no symptoms. In our simulation, we include these people as infectious but not symptomatic. As we employ case confirmation and weighting on the basis of symptoms, these cases are omitted from our analysis: they are counted as successes, rather than fails. They could be confirmed as "cases" if trial participants are routinely PCR tested. Regular PCR testing of all participants would boost power for all trial designs, but might not be operationally feasible.

5.1 Design and analysis choices

As vaccine trials risk being underpowered, due to rarity of events and due to difficulties in recruiting participants, and as ethical response-adaptive design choices come at the cost of power, we explored and presented three design and analysis choices that can be used to mitigate these effects and bolster power. These are the design choices of recruiting contacts at imminent risk of infection, and terminating the trial (or carrying out an interim analysis) once a prespecified number of cases have been confirmed, rather than once a prespecified number of people have been enrolled. The analysis choice is the use of a continuous value to weight cases observed for inclusion in the efficacy calculation and significance test. This analysis choice goes hand in hand with the design choice of ring recruitment, which unavoidably will enrol a number of participants who are already exposed to the disease.

5.1.1 Exclusion criteria implemented at analysis points

We presented a new method for the final analysis, specific to the design of vaccine trial we consider, where it is not clear at randomisation who is eligible for the trial, nor if the vaccination will result in immunity before a vaccinated participant is exposed. Our analysis uses information pertaining to disease and vaccination timings in order to weight confirmed cases in the analysis. The approach is similar to that used by Henao-Restrepo et al. (2017), whose exclusion criterion corresponds to a binary weighting. We use instead a continuous weighting and, in addition, we estimate the vaccine efficacy simultaneously. The effect is an increase in power.

5.1.2 Recruitment schedule: contact tracing

The utility of the ring design for recruitment for vaccine trials for infectious diseases depends most on the balance of predictable to unpredictable edges in infection transmission events (see Table 3). The higher the predictable fraction, the more cases there will be per participant, and the higher the power. In order to power the trial in the first instance (and perhaps to determine if the ring is an appropriate recruitment strategy), an estimate must be made of the balance of predictable to unpredictable transmission events. This might be informed by consideration of data from early efforts of contact tracing: one might ask what proportion of events might have been anticipated, given relationships to prior events, and what sorts of relationships were important. This can be further informed by changes in social structure: e.g. if many predictable events occurred in particular locations, which since have been closed, we can estimate the resulting effect on the predictable-to-unpredictable balance.

Estimation of the predictable fraction could be informed by consideration of available data, e.g. for our example, where we consider household and workplace contacts, we would estimate the predictable fraction as the proportion of new cases that are linked through household and workplace contacts to known, pre-existing cases, or transmission events could be identified following tracing of all contacts (COVID-19 National Emergency Response Center and Epidemiology Case Management Team Korea Centers for Disease Control Prevention, 2020). Alternatively, or additionally, we could use data on contact patterns (Klepac et al., 2020). If we consider that the power associated with our estimated predictable fraction is too low, we could consider adding additional contacts to the network, for example neighbours. If it is beyond the operational capacity of the trial to trace sufficient contacts to achieve a certain desired power, then the ring recruitment strategy is unlikely to deliver an adequately powered trial.

We expect to see a high predictable fraction for diseases whose transmission depends on exchange of or exposure to bodily fluids, such as HIV and EVD, as well as for COVID-19 in societies under "lockdown", where

public spaces are closed and people stay at home, and/or there is extensive quarantining. We expect that the lesser the restrictions to movement and activity, the lesser the fraction of predictable transmission events.

5.1.3 Follow-up time

It can be a challenge with these trial designs to choose appropriate follow-up times. These will be informed by some extent by the parameters governing disease transmission dynamics, and also on the ability to trace contacts, the effective reproduction number, and other quantities that are difficult to know at the outset. Extensive simulation to test a range of values is necessary to narrow down the range of follow-up times that would suit the design objectives.

We use a single follow-up time, 25 days. This was chosen to be large enough to observe cases with confidence that the results were meaningful, in that the case became infected after enrolment and after seroconversion, and small enough so that accumulated data could be used to adapt the trial in a way that delivered on some other objective, e.g. to vaccinate more participants than 50%.

Consistent with all of the above choices would be (a) to follow up all participants beyond the stated number of days: that is, each participant has one endpoint for adaptation and one endpoint final analysis, which could be the outcome at end of the trial. This will increase power for all designs, as more information would accumulate over time. The other option, (b), would be to vaccinate the control arm at the follow-up time. Then the statistical comparison between the two groups terminates at this point (so there is one endpoint only) but there is an ethical benefit of having potentially vaccinated more people. This was the approach used in two EVD trials (Henao-Restrepo et al., 2017; Samai et al., 2018).

5.1.4 How large should the trial be?

We demonstrated the trade-off between defining the size of the trial based on the effective number of cases (resulting in uncertainty in number of participants), and on the effective number of participants enrolled (resulting in uncertainty in power). We advocate choosing the former, as the benefits of being able to anticipate the power exceed the benefits of being able to anticipate the number of participants. Indeed, in outbreaks, the number of participants is often uncertain without it being in the design. We note that the suggestion to end a trial based on the number of cases is proposed also in WHO R&D Blueprint (2020), suggesting that uncertainty in sample size is preferable to uncertainty in power in the context of vaccine trial in an outbreak.

Where there is no vaccine effect, defining trial size based on the total events seen results in a smaller number of participants, and therefore the trial can conclude more quickly, using fewer resources. Where there is a vaccine effect, defining trial size according to number of participants results in more variability in power. The variability in power arises as a result of variability in the number of cases. In some circumstances, it might be that using the number of participants or both the number of cases and the number of participants will be the best way to control power. Because of the dependence on epidemic, disease and vaccination dynamics, the rule to choose, and the components of the rule, become apparent only through simulation.

5.2 Adaptations

5.2.1 Response-adaptive allocation probabilities

We explored four different response-adaptive designs in the context of the vaccine trial. We found that there is a penalty in power of the designs due to the requirement to account for patient drift. The more the allocation deviates from equality, the greater the penalty. Bounding the allocation probabilities between 0.2 and 0.8 protects the designs from very severe penalties, as well as making the designs more realistically acceptable to triallists.

We propose response-adaptive randomisation as a means to achieve some ethical objective, which could be having the most possible treatment successes, aiming to conclude the trial quickly and therefore release an efficacious product for use, or a combination of the two, subject to a particular power. Alternatively, one might consider the balance between power and number of people vaccinated, subject to a fixed trial duration.

For designs that terminate the trial after a prespecified number of cases have been observed (controlling the power), the "cost" to the trial participants is fixed, so the ethics of the design choice is measured in the impact on the rest of the population. This in turn will depend on the local current circumstances of the epidemic. Informally, where the epidemic is growing, maximising vaccination coverage will take priority, as the wider impact of the trial will be to stem the rate of transmission to the outside of the trial population. The cost to power can be met as there will likely be enough participants. Where the epidemic is attenuating, containment is not a primary concern of the trial, but attenuation of trial participants is a risk, so the more ethical choice might be the one that observes the requisite number of cases as fast as possible, in order to license the vaccine for use in other settings. Future work, including modelling a whole population-level epidemic, should quantify this formally, as in Bellan et al. (2017).

We recognise that the two-arm case, with one control arm and one experimental arm, gives limited insight into what a response-adaptive design can offer. In general, in a two-arm trial, one can only increase power (for

a fixed number of participants) at the expense of patient benefit (within the trial) (Williamson et al., 2017; Villar et al., 2015a). We note that even in this case the designs demonstrate a range in behaviours across the spectrum of power vs. patient benefit, even if none exceeds the fixed-randomisation design in this respect. We therefore look forward to multi-arm trials that show the benefits that can be brought in both ethics and efficiency, motivated by the demonstration that the two-arm examples can trade one for the other in different ways.

5.2.2 Early stopping

We demonstrated how existing methods for early stopping can be included in the vaccine trial design: (1) early stopping embedded into the Thompson sampling response-adaptive randomisation scheme, stopping for efficacy when the allocation probability exceeds a threshold; (2) planning a statistical test for efficacy partway through the trial using an α spending strategy to control type 1 error; and (3) stopping for futility at that checkpoint if the number of cases in the experimental arm exceeds a prespecified value.

From this basis, one might try to optimise a particular objective or set of objectives, such as expected number of participants. These considerations could also be informed through connection to the wider epidemic, so that changes to recruitment, including the probability of epidemic extinction, for example, are also included.

5.3 Choice of endpoint

Throughout our analyses we have used a binary endpoint weighted by a retrospective exclusion criterion. A time-to-event–type analysis (TTE) would be enabled by the data we assume is collected. We use a weighted binary outcome so that the temporal element of our outcome is associated exclusively with the exclusion criterion. By weighting a binary outcome we can account more naturally for "noise" in the data, such as time from infection to symptom onset, as well as exclusion, than would be possible in a TTE analysis.

Additionally, we note that a central assumption of a TTE analysis is that of proportional hazards: that the hazard function for any two individuals is proportional at any moment in time and the implied ratio does not vary with time. With a network-based trial, we can verify that this assumption is violated almost certainly: a person's hazard depends on the infection status of their neighbours. The hazard ratio between two people would then invert if one person has an infectious contact soon after their enrolment but not at the end of their follow-up time (e.g. if they are a contact of the index case), and the other person had no infectious contacts soon after enrolment but did at the end of their follow-up time (e.g. if they are a contact of a contact, and their contact becomes symptomatic during the trial).

We suggest that TTE analysis would be enabled for a trial such as this one through parametrisation by exposure data, quantified as the cumulative time spent with a contact (or contacts) in the infectious state. This might be more easily quantified for a disease such as EVD, where symptoms correspond to risk of infection and so knowledge of the duration of symptoms of a person's contacts enables quantification of their accumulated risk. Phone apps developed for COVID-19 would provide similar data towards parametrising such a model.

5.4 Vaccine efficacy estimates

In our simulations we consistently underestimate the vaccine effect. There is an apparent trend that the higher the power, the closer the VE estimate to the true effect of 0.7, which we would expect as higher power implies more trials in which efficacy is concluded, and where efficacy is concluded the VE estimate is likely to be higher. We accounted for the the recruitment design in our estimates, which allows for people vaccinated to have been infected before vaccination, or before seroconversion. Thus we are making an isolated estimate of the vaccine efficacy after seroconversion.

Our simulation model could be used also to investigate the impact of the vaccination strategy, in addition to the effect of the vaccine. Such a comparison would use a different participant weighting to isolate the effect of interest, for example having all weights equal to 1 would indicate the effect of the overall intervention on the whole eligible population. Similarly, using the same model with cluster randomisation would permit investigation of effects of herd immunity, again on the whole eligible population.

5.5 Relation to the epidemic

We have simulated a trial to operate within an epidemic. However, we have not simulated that epidemic itself, and so we have not included in our analyses quantification of the immediate impact of the trial on the wider population. So, while we have compared measures of patient benefit, these have been limited to those in the trial, rather than those outside. We considered the prevented export infections as a means to compare different designs, which might form a starting point for assessing the wider impact of the trial on the epidemic. However, we defer answering that question directly to later work.

Explicit inclusion of the trial in an epidemic would also permit a realistic assessment of the impact of heterogeneity in disease state of participants, of contact networks that interact or intersect, and realistic time

trends. Then the likelihood of running out of eligible trial participants could be assessed and included among criteria for trial design assessment, and different paces of trials could be considered: we supposed that one contact network was initiated per day. In an epidemic, there might be the possibility to recruit multiple contact networks per day, which would impact on the characteristics of the adaptive designs.

References

Bellan, S. E., Pulliam, J. R. C., Van Der Graaf, R., Fox, S. J., Dushoff, J., and Meyers, L. A. (2017). Quantifying ethical tradeoffs for vaccine efficacy trials during severe epidemics. *bioRxiv*.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, 25(2):290–302.

Brueckner, M., Titman, A., Jaki, T., Rojek, A., and Horby, P. (2018). Performance of different clinical trial designs to evaluate treatments during an epidemic. *PLoS ONE*, 13(9):e0203387.

Camacho, A., Eggo, R. M., Funk, S., Watson, C. H., Kucharski, A. J., and Edmunds, W. J. (2015). Estimating the probability of demonstrating vaccine efficacy in the declining Ebola epidemic: A Bayesian modelling approach. *BMJ Open*, 5(12):1–6.

COVID-19 National Emergency Response Center and Epidemiology Case Management Team Korea Centers for Disease Control Prevention (2020). Contact transmission of COVID-19 in South Korea: Novel investigation techniques for tracing contacts. *Osong Public Health and Research Perspectives*, 11(1):60–63.

Danon, L., Brooks-Pollock, E., Bailey, M., and Keeling, M. J. (2020). A spatial model of CoVID-19 transmission in England and Wales: early spread and peak timing. *medRxiv*, page 2020.02.12.20022566.

Dean, N. E., Gsell, P.-S., Brookmeyer, R., De Gruttola, V., Donnelly, C. A., Halloran, M. E., Jasseh, M., Nason, M., Riveros, X., Watson, C. H., Henao-Restrepo, A. M., and Longini, I. M. (2019). Design of vaccine efficacy trials during public health emergencies. *Science Translational Medicine*, 11(499):eaat0360.

Ebola ça Suffit Ring Vaccination Trial Consortium (2015). The ring vaccination trial: a novel cluster randomised controlled trial design to evaluate vaccine efficacy and effectiveness during outbreaks, with special reference to Ebola. *BMJ (Clinical research ed.)*, 351(July):h3740.

ECDC (2020). Resource estimation for contact tracing, quarantine and monitoring activities for COVID-19 cases in the EU/EEA. Technical Report March, European Centre For Disease Prevention And Control.

Fenner, F., Henderson, D. A., Arita, I., Jezek, Z., and Ladnyi, I. D. (1988). *Smallpox and its eradication*, volume 6.

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L. C., van Elsland, S., Thompson, H., Verity, R., Volz, E., Wang, H., Wang, Y., Walker, P. G., Walters, C., Winskill, P., Whittaker, C., Donnelly, C. A., Riley, S., and Ghani, A. C. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID- 19 mortality and healthcare demand. Technical Report March, Imperial College COVID-19 Response Team.

Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., and Fraser, C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 6936(March):eabb6936.

Friede, T. and Kieser, M. (2002). On the inappropriateness of an EM algorithm based procedure for blinded sample size re-estimation. *Statistics in Medicine*, 21(2):165–176.

Fyles, M., Fearon, E., and Working, M. C. (2020). Household structured contact tracing: Branching process model. Technical report.

Gould, A. L. and Shih, W. J. (1992). Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. *Communications in Statistics - Theory and Methods*, 21(10):2833–2853.

He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B. J., Li, F., and Leung, G. M. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(May).

Henao-Restrepo, A. M., Camacho, A., Longini, I. M., Watson, C. H., Edmunds, W. J., Egger, M., Carroll, M. W., Dean, N. E., Diatta, I., Doumbia, M., Draguez, B., Duraffour, S., Enwere, G., Grais, R., Gunther, S., Gsell, P. S., Hossmann, S., Watle, S. V., Kondé, M. K., Kéïta, S., Kone, S., Kuisma, E., Levine, M. M., Mandal, S., Mauget, T., Norheim, G., Riveros, X., Soumah, A., Trelle, S., Vicari, A. S., Røttingen, J. A., and Kieny, M. P. (2017). Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease:

final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!). *The Lancet*, 389(10068):505–518.

Henao-Restrepo, A. M., Longini, I. M., Egger, M., Dean, N. E., Edmunds, W. J., Camacho, A., Carroll, M. W., Doumbia, M., Draguez, B., Duraffour, S., Enwere, G., Grais, R., Gunther, S., Hossmann, S., Kondé, M. K., Kone, S., Kuisma, E., Levine, M. M., Mandal, S., Norheim, G., Riveros, X., Soumah, A., Trelle, S., Vicari, A. S., Watson, C. H., Kéïta, S., Kieny, M. P., and Røttingen, J. A. (2015). Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial. *The Lancet*, 386(9996):857–866.

Hitchings, M. D., Lipsitch, M., Wang, R., and Bellan, S. E. (2018). Competing effects of indirect protection and clustering on the power of cluster-randomized controlled vaccine trials. *American Journal of Epidemiology*, 187(8):1763–1771.

Hitchings, M. D. T., Grais, R. F., and Lipsitch, M. (2017). Using simulation to aid trial design: Ring-vaccination trials. *PLoS Neglected Tropical Diseases*, 11(3):1–12.

Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. John Wiley Sons, Ltd., New Jersey.

Huang, L., Bai, J., Yu, H., and Chen, F. (2018). Sample size re-estimation without un-blinding for time-to-event outcomes in oncology clinical trials. *Journal of Biomedical Research*, 32(1):23–29.

Kahn, R., Rid, A., Smith, P. G., Eyal, N., and Lipsitch, M. (2018). Choices in vaccine trial design in epidemics of emerging infections. *PLoS Medicine*, 15(8):1–12.

Kahn, R., Villar, S. S., and Lipsitch, M. (2020). Innovative Vaccine Trial Designs for EID Outbreak Response. In preparation.

Kennedy, S. B., Bolay, F., Kieh, M., Grandits, G., Badio, M., Ballou, R., Eckes, R., Feinberg, M., Follmann, D., Grund, B., Gupta, S., Hensley, L., Higgs, E., Janosko, K., Johnson, M., Kateh, F., Logue, J., Marchand, J., Monath, T., Nason, M., Nyenswah, T., Roman, F., Stavale, E., Wolfson, J., Neaton, J. D., and Lane, H. C. (2017). Phase 2 placebo-controlled trial of two vaccines to prevent ebola in Liberia. *New England Journal of Medicine*, 377(15):1438–1447.

Kennedy, S. B., Neaton, J. D., Lane, H. C., Kieh, M. W., Massaquoi, M. B., Touchette, N. A., Nason, M. C., Follmann, D. A., Boley, F. K., Johnson, M. P., Larson, G., Kateh, F. N., and Nyenswah, T. G. (2016). Implementation of an Ebola virus disease vaccine clinical trial during the Ebola epidemic in Liberia: Design, procedures, and challenges. *Clinical Trials*, 13(1):49–56.

Kiss, I. Z., Miller, J., and Simon, P. (2017). *Mathematics of Epidemics on Networks From Exact to Approximate Models*. Springer.

Klepac, P., Kucharski, A. J., Conlan, A. J., Kissler, S., Tang, M., Fry, H., and Gog, J. R. (2020). Contacts in context: large-scale setting-specific social mixing matrices from the BBC Pandemic project. *medRxiv*, page 2020.02.16.20023754.

Kucharski, A. J., Klepac, P., Conlan, A., Kissler, S. M., Tang, M., Fry, H., Gog, J., and Edmunds, J. (2020). Effectiveness of isolation, testing, contact tracing and physical distancing on reducing transmission of SARS-CoV-2 in different settings. *medRxiv*, page 2020.04.23.20077024.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S., Lau, E. H., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T. T., Wu, J. T., Gao, G. F., Cowling, B. J., Yang, B., Leung, G. M., and Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, pages 1–9.

Long, Q.-X., Liu, B.-Z., Deng, H.-J., Wu, G.-C., Deng, K., Chen, Y.-K., Liao, P., Qiu, J.-F., Lin, Y., Cai, X.-F., Wang, D.-Q., Hu, Y., Ren, J.-H., Tang, N., Xu, Y.-Y., Yu, L.-H., Mo, Z., Gong, F., Zhang, X.-L., Tian, W.-G., Hu, L., Zhang, X.-X., Xiang, J.-L., Du, H.-X., Liu, H.-W., Lang, C.-H., Luo, X.-H., Wu, S.-B., Cui, X.-P., Zhou, Z., Zhu, M.-M., Wang, J., Xue, C.-J., Li, X.-F., Wang, L., Li, Z.-J., Wang, K., Niu, C.-C., Yang, Q.-J., Tang, X.-J., Zhang, Y., Liu, X.-M., Li, J.-J., Zhang, D.-C., Zhang, F., Liu, P., Yuan, J., Li, Q., Hu, J.-L., Chen, J., and Huang, A.-L. (2020). Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nature Medicine*, pages 1–4.

Marzi, A., Robertson, S. J., Haddock, E., Feldmann, F., Hanley, P. W., Scott, D. P., Strong, J. E., Kobinger, G., Best, S. M., and Feldmann, H. (2015). VSV-EBOV rapidly protects macaques against infection with the 2014/15 Ebola virus outbreak strain. *Science*, 349(6249):739–742.

Mehta, C., Gao, P., Bhatt, D. L., Harrington, R. A., Skerjanec, S., and Ware, J. H. (2009). Optimizing trial design. Sequential, adaptive, and enrichment strategies. *Circulation*, 119(4):597–605.

Nason, M. (2016). Statistics and logistics: Design of Ebola vaccine trials in West Africa. *Clinical Trials*, 13(1):87–91.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2):191–199.

Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C. N., and Ricks, M. L. (2001). Optimal Adaptive Designs for Binary Response Trials With Three Treatments. *Biometrics*, 57:909–913.

Samai, M., Seward, J. F., Goldstein, S. T., Mahon, B. E., Lisk, D. R., Widdowson, M. A., Jalloh, M. I., Schrag, S. J., Idriss, A., Carter, R. J., Dawson, P., Kargbo, S. A., Leigh, B., Bawoh, M., Legardy-Williams, J., Deen, G., Carr, W., Callis, A., Lindblad, R., Russell, J. B., Petrie, C. R., Fombah, A. E., Kargbo, B., McDonald, W., Jarrett, O. D., Walker, R. E., Gargiullo, P., Bash-Taqi, D., Gibson, L., Fofanah, A. B., Schuchat, A., Neuzil, K., Insip, H., Risi, G., and Sow, S. (2018). The Sierra Leone Trial to Introduce a Vaccine Against Ebola: An Evaluation of rVSVΔG-ZEBOV-GP Vaccine Tolerability and Safety during the West Africa Ebola Outbreak. *Journal of Infectious Diseases*, 217(December 2014):S6–S15.

Schoenfeld, D. A. . (1983). Sample-size formula for the proportional-hazards regression model. *International Biometric Society*, 39(2):499–503.

Simon, R. and Simon, N. R. (2011). Using randomization tests to preserve type I error with response-adaptive and covariate-adaptive randomization. *Statistics Probability Letters*, 81(7):767–772.

Tapiwa, G., Cécile, K., Dongxuan, C., Andrea, T., Christel, F., Jacco, W., and Niel, H. (2020). Estimating the generation interval for COVID-19 based on symptom onset data. *medRxiv*, pages 1–13.

Teel, C., Park, T., and Sampson, A. R. (2015). EM estimation for finite mixture models with known mixture component size. *Commun Stat Simul Comput*, 44(6):1545–1556.

Villar, S. S., Bowden, J., and Wason, J. (2015a). Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statistical Science*, 30(2):199–215.

Villar, S. S., Bowden, J., and Wason, J. (2018). Response-adaptive designs for binary responses: How to offer patient benefit while being robust to time trends? *Pharmaceutical Statistics*, 17(2):182–197.

Villar, S. S., Wason, J., and Bowden, J. (2015b). Response-adaptive randomization for multi-arm clinical trials using the forward looking Gittins index rule. *Biometrics*, 71(4):969–978.

Wason, J. M. S., Brocklehurst, P., and Yap, C. (2019). When to keep it simple – adaptive designs are not always useful. *BMC Medicine*, 17(1):1–7.

WHO Disease Surveillance and Response Programme Area Disease Prevention and Control Cluster (2014). Contact tracing during an outbreak of Ebola virus disease. Technical Report September.

WHO R&D Blueprint (2020). An international randomised trial of candidate vaccines against COVID-19. Technical report.

Widdowson, M.-A., Schrag, S. J., Carter, R. J., Carr, W., Legardy-Williams, J., Gibson, L., Lisk, D. R., Jalloh, M. I., Bash-Taqi, D. A., Kargbo, S. A. S., Idriss, A., Deen, G. F., Russell, J. B., McDonald, W., Albert, A. P., Basket, M., Callis, A., Carter, V. M., Ogunsanya, K. R. C., Gee, J., Pinner, R., Mahon, B. E., Goldstein, S. T., Seward, J. F., Samai, M., and Schuchat, A. (2016). Implementing an Ebola Vaccine Study — Sierra Leone. *MMWR Supplements*, 65(3):98–106.

Williamson, S. F., Jacko, P., Villar, S. S., and Jaki, T. (2017). Europe PMC Funders Group Europe PMC Funders Author Manuscripts A Bayesian adaptive design for clinical trials in rare diseases. pages 136–153.

A The COVID-19 model

There are two components to the model: the network that describes relationships between individuals who have the potential to be involved in the trial; and the transmission model that describes the dynamics of disease over the network.

The relationships in the network define the contact structure, and facilitate infection transmission. We make the assumption that our network for disease propagation includes (as a subnetwork) the network for contact tracing. The disease transmission tree is also a subnetwork of the whole network, and might include some edges not in the (relationship) contact network. In terms of modelling, these structures would need to change for a very large population: we consider here only populations of the order of one thousand people.

Both the network model and the disease transition model will be specific to a particular setting and, indeed, might change or become better informed over the course of the trial. Therefore, we use models without supposing that they will reflect the "truth" for any particular disease or scenario. Instead, we make choices we believe to be plausible in order to demonstrate the general methods, whose features will persist for settings such as these within the parameters we have described, i.e., diseases that pass from person to person, through contacts that can be traced.

A.1 Network model

We construct a network to represent relationships between individuals, where individuals are the nodes (or vertices) of the network, and relationships are edges between nodes. Infection is transmitted from one node to another across an edge that joins them. Therefore we want the network to include all possible transmissive relationships. These include interpersonal relationships, as well as random relationships, as it is likely that COVID-19 spreads between people who encounter each other only transiently as well as between people who know each other.

In our simulation, we consider a society in a moment where "social distancing" (or "physical distancing") is insisted upon: schools are closed but employees go to work; social spaces such as hospitality and entertainment establishments are closed. Therefore we consider three types of relationships: within the home, in the workplace, and transient (or random). Transient relationships include encounters when in transit, or in shops, or due to the sharing of infrastructure. This is slightly different from the relationships considered in Kucharski et al. (2020), where interactions were localised to places (such as home or school or work) and within those spaces, a proportion of encounters are with people who are "acquaintances", that is, people who have been met before, and a proportion of encounters will be with new people. In contrast, we designate that all work and home contacts are "acquaintances" and all random relationships are not, by definition.

For our simulation, we ascribe to individuals ages: below 19, 19 to 65, and over 65. We make this distinction as there are data available to guide the construction of the residency part of the network. People live in dwellings according to the UK census (`https://www.ons.gov.uk/peoplepopulationandcommunity/housing/adhocs/008634ct08192011censushouseholdtypehouseholdsizeandageofusualresidentshouseholdsenglandandwales`). We choose to simulate 500 dwellings in total. People aged 19 to 65, and one fifth of people aged 65+, are connected to approximately 15 other people via a "workplace". The number of people in the workplace is reflective of the likely number of people with whom an infrastructure is shared, rather than the number of colleagues. Together, the household edges and the workplace edges are the predictable edges that make up the contact network (the network of acquaintances, in the terminology of Kucharski et al. (2020)). Finally, we include approximately 10 random connections per person to any other person. These are unpredictable (random/transient) edges.

There are many things we have omitted from this example, which might be important to include in a particular application. Omissions include: institutional residences, including care homes, halls of residence, barracks and prisons; structured contacts between children (such as schools); structured social contacts; extended family contacts; neighbour contacts; social spaces. The "workplace" element in our simulation lacks nuances such as details of which people in the population commute to a regular workplace, and with whom this workplace is shared. Such details could be filled in using information from the census and the Labour Force Survey for the UK.

It is also worth noting that these network structures might change, e.g. in response to the progress of the epidemic. These changes will affect the extent to which traced contacts account for new observed cases. If the fraction is smaller than anticipated, the study will be underpowered, as fewer events than expected will be observed in the trial population. To maintain the trial's power, the key is to establish who the index cases shared space with in the last week and who these contacts will likely share space with in the coming week. Hopefully, then, it will still be possible to capture individuals at imminent risk of exposure in a natural way.

A.1.1 Definitions

Our network is an undirected graph,

$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

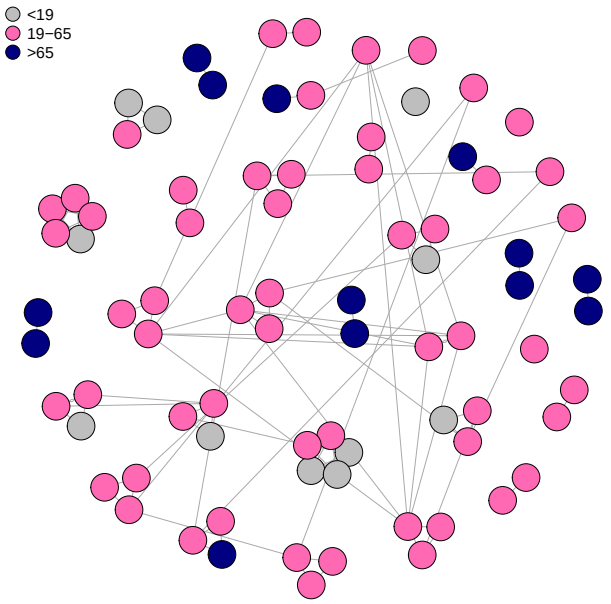

Figure 2: Example contact network showing 80 people in a contact network. "Predictable" edges between housemates and colleagues are shown. Individuals are coloured by age group and clustered into households.

with vertices, or nodes,

$$\mathcal{V} = \{i, i = 1, ..., N_I\}$$

representing the N_I individuals, which we index with i, and edges

$$\mathcal{E} \subseteq \{\{i, j\} | (i, j) \in \mathcal{V}^2 \cap i \neq j\}$$

representing connections, or relationships, between individuals. Each node i has a set of attributes $\{a_1(i), a_2(i), ..., a_{N_a}(i)\}$.

The network operates at the level of the individuals. Each individual belongs to a household $\mathcal{H}_h$ and every individual in a household is connected to every other individual in the household, as in Fyles et al. (2020), such that the induced subgraph $\mathcal{G}_1[\mathcal{H}_h]$ is completely connected:

$$\mathcal{E}_1[\mathcal{H}_h] = \{\{i, j\} | (i, j) \in \mathcal{V}[\mathcal{H}_h]^2 \cap i \neq j\}.$$

Let $\mathcal{H} = \{\mathcal{H}_h, h = 1, ..., N_H\}$ be the set of N_H households, which forms a partition of the vertices $\mathcal{V}$ (all subsets are mutually disjoint and their union is equal to the set). Let $a_1(i)$ be the household of individual i, i.e. $a_1(i) = h \Leftrightarrow i \in \mathcal{V}[\mathcal{H}_h]$. Then the size of a household is

$$\mathcal{O}(\mathcal{H}_h) = \sum_{i=1}^{N_I} \mathbf{1}_{a_1(i)=h}.$$

We start with $N_H = 500$ households with $\mathcal{O}(\mathcal{H}_h)$ people in each. The age and number distribution follows `https://www.ons.gov.uk/peoplepopulationandcommunity/housing/adhocs/008634ct08192011censushouseholdtypehouseholdsi` That is, each person has an age attribute,

$$a_2(i) \in \{1, 2, 3\}$$

where $a_2(i) = 1$ if person i is under 19, $a_2(i) = 2$ if person i is 19 to 65, and $a_2(i) = 3$ if person i is over 65.

We define an individual i to be part of the workforce via a "worker" variable $a_3(i)$, where

$$\begin{aligned}
a_2(i) = 1 &\implies a_3(i) = 0, \\
a_2(i) = 2 &\implies a_3(i) = 1, \\
a_2(i) = 3 &\implies \begin{cases} a_3(i) = 0 & \text{with probability } 0.8 \\ a_3(i) = 1 & \text{with probability } 0.2. \end{cases}
\end{aligned}$$

There are $A_3 = \sum_i \mathbf{1}_{a_3(i)=1} \approx 731$ people in the workforce. We define $A_3/15 \approx 49$ workplaces to which people with $a_3(i) = 1$ (that is, all those aged 19 to 65, and 1/5 of people over 65) are assigned following a

multinomial distribution. We choose to place on average 15 people in a "workplace", $\mathcal{W}_w$, which represents not their employment structure but a close shared use of the infrastructure. A useful guide might be how many toilets per person a place of work should have.

The workplaces are completely connected:

$$\mathcal{E}[\mathcal{W}_w] = \{\{i,j\}|(i,j) \in \mathcal{V}[\mathcal{W}_w]^2 \cap i \neq j\}.$$

Finally, 1000 totally random edges are added, which amounts to around ten per person:

$$\Pr\left(\{i,j\} \in \mathcal{E}\right) = \frac{10}{\left(\sum_h \mathcal{O}(\mathcal{H}_h)\right)}.$$

These edges correspond to potential transmission encounters that would not be recalled or anticipated through contact tracing. The result is an average of 20 connections per person, of which ten have a weight of 1 and ten have a weight of 0.1.

A.1.2 Parametrisation

The parameters we have used and their provenance are listed in Table 9.

Table 9: Parameters used in the COVID-19 disease transmission and vaccine trial model. $\mathcal{N}(l, \mu, \sigma)$ denotes a normal distribution truncated at l.

Parameter	Value	Source
Number of households	500	Chosen with reference to other choices
Household size	Draws from raw data	UK 2011 census
Workplace size	$\sim$ Poisson(15)	`https://www.hse.gov.uk/contact/faqs/toilets.htm`
β (per-contact infection or transmission rate)	0.01	Chosen with reference to other choices
Predictable edge weight (χ_h, χ_w)	1	Chosen with reference to other choices
Unpredictable edge weight (χ_n)	0.1	Chosen with reference to other choices
Incubation period	$\sim 2 + \Gamma(\text{shape} = 13.3, \text{rate} = 4.16)$	Li et al. (2020)
Infectious period	$\sim 1 + \Gamma(\text{shape} = 1.43, \text{rate} = 0.549)$	Li et al. (2020)
Time to enrol whole contact network	$\sim \mathcal{N}(l = 0, 10.32, 4.79)$	From Henao-Restrepo et al. (2017). One could use ECDC (2020). Fyles et al. (2020) use Poisson Parameter $\sim$ Uniform(1.5, 2.5)
Time to vaccine-induced seroconversion	$\sim \Gamma(\text{shape} = 3, \text{rate} = 1)$	Plausibly all seroconverted within 14 days

A.2 Disease and trial state transitions

Individuals' disease states and possible transitions are described by a compartmental model (Figure 3), as in Camacho et al. (2015). Possible state transitions are disease progression ($S \to E \to I_1 \to I_2 \to R$, where S denotes Susceptible, E exposed, I infectious, and R removed), similarly to Danon et al. (2020), and trial enrolment. There are rules for both types of transitions. Transition rates depend on the rules and the states of the neighbours.

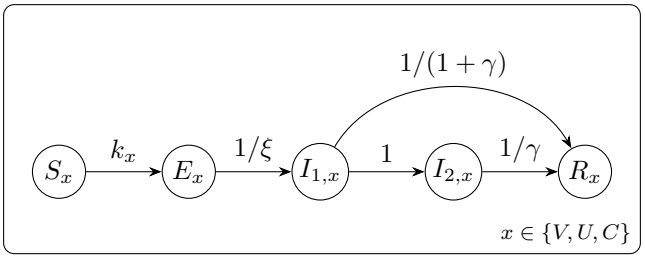

Figure 3: Disease-state transition model for members of the population who aren't enrolled (U) and those enrolled and vaccinated (V) and those enrolled to the control group (C). Arrows show possible transitions between states, labelled by the rates. Everyone starts in State S_U, except for one person who starts in E_U. Rates of transition of person i depend on (a) the current state of person i and (b) the current states of the neighbourhood of person i. The instantaneous rate of infection (or infection hazard), k_x, depends on vaccination status x. People spend ξ days in the exposed state before moving to the infectious state. In the first phase of infectiousness, I_1, all people are infectious and asymptomatic for one day (Kucharski et al., 2020). 20% of people remain asymptomatic for a further infectious period of γ days. The other 80% are symptomatic for γ days.

Our simulated trial operates on a time unit of one day, and has total duration of the order of 100 days. For simplicity, we assume that we begin initiation of one contact network (corresponding to one index case) every day. Initiation is the moment where all nodes in the network are in state S_U except one who is on their first day as state E_U. Enrolment begins when this node reaches state I_2.

Each participant who is enrolled has as their reference day the day on which they were enrolled. Therefore every participant has their own individual timeline which can be referenced to others by the day of their enrolment.

A.2.1 Disease transmission rules

A person in the E state becomes infectious, I_1, 24 hours before entering the I_2 state as in Kucharski et al. (2020). They are infectious but not symptomatic. For our simulation, we assume state/society rules that symptomatic persons are to shelter in place (which means, for our simulation, they do not leave their homes). Therefore, as an infectious I_1, a person can infect their home contacts, their work contacts and their random contacts. The majority of I_1 people become I_2 people. As an I_2, a person can infect only their home contacts. Some infectious people, however, are not symptomatic, so these people behave as I_1s, in terms of who they infect. This is represented by an additional edge from I_1 to R, bypassing I_2, where the person stays in the I_1 state for an addition period, equal to the period they would have spent as an I_2 if symptomatic. Transitions to state R imply removal from the infectious population: this can be due to recovery or hospitalisation, and hence isolation.

A susceptible individual i belonging to arm x becomes "exposed" (i.e. transitions to state E) with rate

$$k_x(i) = \beta \left(\chi_h \mathcal{M}_i^{(h)} + \chi_h \mathcal{L}_i^{(h)} + \chi_w \mathcal{L}_i^{(w)} + \chi_n \mathcal{L}_i^{(n)} \right) (1 - \eta \cdot x_i), \quad x \in \{V, U, C\},$$

where h refers to household, w to workplace, and n to random. $\chi_h > 0$ is the scalar (or edge weight) for household contacts, $\chi_w > 0$ is the scalar (or edge weight) for workplace contacts, and $\chi_n > 0$ is the scalar (or edge weight) for random contacts. One could specify a different χ_h for pre- and post-symptomatic infectiousness, or even a changing profile over time, as described in He et al. (2020). $x_i = 1$ if person i is vaccinated and 0 otherwise. $\eta \leq 1$ is the vaccine efficacy, and $\beta > 0$ is the per-contact rate at which infectious people infect their susceptible contacts.[5] $\mathcal{M}_i^{(h)}$ is the number of I_2 household (h) contacts of susceptible i.

$$\mathcal{L}_i^{(y)} = \sum_{j \in \mathcal{N}_{(y)}(i)} \mathbf{1}_{j \in \{I_{1,\cdot}\}}, \quad y \in \{h, w, n\}$$

is similarly defined as the number of contacts who are infectious but not symptomatic.

[5]The unit of β is per contact per day; the scalars χ are unitless. Thought of as edge weights, they transform the static (or quenched) network into a dynamic or adaptive network through link deactivation (Kiss et al., 2017).

Note that there is no infection other than from a contact (with the exception of the simulated "time trends", Figure 5). Thus any infection within a contact network came (directly or indirectly) from the contact network's index case.

Our model yields new infection events as occurring from pre-symptomatic people 58% of the time and from symptomatic people 42% of the time, omitting transmissions from people who never become symptomatic. In comparison to reported estimates (44% from 77 recorded transmission pairs (He et al., 2020), and 48 and 62% in Singapore and Tianjin, respectively (Tapiwa et al., 2020), we confirm that our simulation scenario is consistent with one with quick quarantine of close contacts (He et al., 2020).

A.2.2 Trial rules

Our simulation mimics the trial rules of the ring recruitment strategy (Henao-Restrepo et al., 2015), with the exception that we randomise at the individual, rather than cluster, level, and therefore do not exclude individuals on the basis of contact-network overlap with existing enrolled networks.

In our trial we define contact tracing as identifying only existing relationships ("acquaintances", or "those met before", in the terminology of Kucharski et al. (2020)) that might be, or might have been, a means for transmission. That is, a newly diagnosed person is asked to recall all of their contacts, and these individuals are contacted and asked likewise to list all their contacts. This makes our simulation and trial design most like the "self-isolation and manual contact tracing of acquaintances" of Kucharski et al. (2020). The relationships that are of interest will depend on the society and any concurrent actions, guidance or instruction from the state, which in our simulation include home and workplace contacts, while random contacts remain unpredictable. We assume unpredictable relationships are not recalled in contact tracing and, similarly, cannot be anticipated for recruitment of trial participants. We assume work and home relationships are recalled perfectly. By recruiting "contacts" and "contacts of contacts" into the trial (the "two-step tracing" of Fyles et al. (2020)), we are recruiting the people the index case lives with and their colleagues, and those the case works with and their housemates.

For our purposes, an index case for a contact network is a person identified as being in state $I_{2,\cdot}$ after the initiation of the trial. Eligible people are traced as described in ECDC (2020) and those who consent are enrolled as soon as they are identified and give their consent. Susceptible and exposed people are eligible for enrolment if they are a "predictable" contact of the index case and they are not already enrolled in the trial. Symptomatic $I_{2,U}$ people are excluded on the basis of their symptoms. R_U are excluded on the basis of their history, which we assume a perfect knowledge of in our simulation. This could result either from people being able to identify having had COVID-19, or from there being an accurate and reliable antibody test. Inclusion of R_U people in the simulated trial would result in a dilution of infections and would therefore require enrolment of more participants to maintain power. Their inclusion is advocated for reasons of safety testing (`https://www.who.int/publications/i/item/an-international-randomised-trial-of-candidate-vaccines-against-covid-19`, `https://clinicaltrials.gov/ct2/show/NCT04405076`). Elsewhere seropositive people are excluded from vaccine trials (`https://www.clinicaltrialsregister.eu/ctr-search/trial/2020-001228-32/GB`).

The enrolment rate is $0 \leq \epsilon \leq 1$, which is the probability for each eligible person to enrol, where we use only contact structure and lack of symptoms to define eligibility. We use $\epsilon = 0.7$.[6] Enrolled participants are randomised to the experimental arm with probability $0 \leq \pi_1 \leq 1$, and $\pi_1 + \pi_0 = 1$. Enrolment takes time δ. Those vaccinated have an additional wait time before reaching state S_V, which is development of immunity (or time to seroconversion), and which takes time τ. Transition from S_U to E_U is possible in this wait time. We assume that the vaccine effect before seroconversion is zero and that the vaccine effect after is the full effect of the vaccine.

Individuals who are in state E and are unenrolled are enrolled with the same probability (ϵ) as the susceptibles S_U, as they are asymptomatic. The same wait time for enrolment applies, but time to seroconversion does not, as the individual is already infected. If the participant transitions to $I_{2,U}$ before the recruitment time elapses, they will be excluded from the trial, as they will be showing symptoms and can be tested for confirmation.

Result accrual relies on surveillance and self reporting. For our simulations we assume that a fraction 0.2 of infectious individuals are asymptomatic. These infection events go unreported in the trial results (but still contribute to onward transmission).

We simulate one contact network at a time, beginning when the index case is identified. Transmission in each contact network is independent of all other contact networks in the trial. The contact networks are related only through the time reference, in that one contact network is initiated on each day. Ideally, one would simulate a whole population that encompasses potentially thousands of distinct contact networks, so that each contact network might have a history that is affected by the events stemming from other contacts, and the moments for contact-network enrolment would arise in a more natural way. Here, however, we present only an

[6] Henao-Restrepo et al. (2017) report that 50% of people identified were eligible and enrolled, where their eligibility criteria excluded people who were pregnant, breastfeeding, or under the age of 18.

approximation to that population, and omit entirely any shared history, save the accrued results that determine the response-adaptive allocation probabilities in the adaptive designs.

Finally, note that, for simplicity, we assume that randomisation, enrolment and vaccination are all assumed to happen on a single day for each individual, although the day will differ between individuals.

B Analyses, and exclusion criterion implemented at analysis points

Here we detail all equations to accompany Section 2.1, which describes the different ways we explore to analyse the outcome. We present the methods in the same order and use a single framework that describes all the methods in the same way.

B.1 Analysis of raw data

We have $j = 1, ..., N_I$ individuals. Each has a vaccination status, x_j, and a disease status y_j, where the vaccination status is dictated by the trial design and the disease status from the underlying epidemic model:

$$x_j = \begin{cases} 0 & \text{person } j \text{ not vaccinated} \\ 1 & \text{person } j \text{ vaccinated} \end{cases}$$

and

$$y_j = \begin{cases} 0 & \text{person } j \text{ not diagnosed} \\ 1 & \text{person } j \text{ diagnosed} \end{cases}$$

at the end of the trial.

The test statistic is a standard normal variable Z, where

$$Z = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{\sigma_0 + \sigma_1}}, \tag{1}$$

$$\hat{p}_v = \frac{\sum_{j:x_j=v, y_j=0} \omega_j}{N_v}, \tag{2}$$

$$N_v = \sum_{j:x_j=v} \omega_j \tag{3}$$

$$\omega_j = 1 \quad \forall j,$$

and

$$\sigma_v = \frac{\hat{p}_v(1-\hat{p}_v)}{N_v}. \tag{4}$$

p_v is the true probability of not being a confirmed case if in arm v, and $\hat{p}_v$ is our estimate of it, defined as the proportion of people in arm v not confirmed, where $v = 0$ is the control arm and $v = 1$ is the experimental arm; N_v is the total number in arm v; and σ_v is the variance of the estimator $\hat{p}_v$. Power is defined as the proportion of Z values that exceed 1.64, which is the 95th quantile of a standard normal distribution. ω_j is the weight of person j which, for the unweighted method, is 1 for all participants. In the descriptions that follow, we see that the weights ω define the retrospective exclusion criterion so that all methods use the same calculation and each is defined only by the definition of the weights.

B.2 Analysis using binary weighting

The binary-weighting method proceeds as above but considers also the day of commencement of symptoms, s_j. Individuals are excluded if $s_j < 9$ relative to a randomisation day of 0. We write this as the weight, ω_j, for each individual j, so that a weight of 0 equates to exclusion:

$$\omega_j = \begin{cases} 0 & s_j < 9 \\ 1 & s_j \geq 9 \text{ or } y_j = 0 \end{cases}$$

These are used together with Equations 1, 2, 3, and 4 as before.

B.3 Analysis using continuous weighting

Given person j's symptoms began on day s_j relative to their randomisation day of 0, τ and ξ are their unknown time to seroconversion[7] and incubation time, respectively. The probability they were infected after the trial began is $P(\tau + \xi < s_j)$. We assume τ and ξ are distributed between individuals as $\Gamma(\text{shape} = 3, \text{rate} = 1)$ and $2 + \Gamma(\text{shape} = 13.3, \text{rate} = 4.16)$. Hence the distribution of $\tau + \xi$ is estimated by matching moments using the Welch–Satterthwaite equation, as described in Box (1954).

[7]Note that we include time to seroconversion also for the control group, who don't receive the vaccine, and don't seroconvert.

We estimate the vaccine efficacy $0 \leq \hat{\eta} \leq 1$ as

$$\hat{\eta} = 1 - \frac{f_1}{N_1} \Big/ \frac{f_0}{N_0},$$

where

$$\hat{f}_v = \sum_{j:x_j=v, y_j=1} \omega_j.$$

Suppose from the cumulative distribution function of the gamma distribution we have a nominal probability, i.e. neglecting the effect of the vaccine, q_j that the day person j was infected, d_j, was after seroconversion on day D_c. We write the complement, the probability that person j was infected before day D_c, as $r_j = 1 - q_j$. We re-estimate their probability given that person j was vaccinated ($x_j = 1$):

$$\Pr(d_j > D_c | x_j = 1) = \frac{(1-\hat{\eta})x_j}{y_j + (1-\hat{\eta})x_j}, \tag{5}$$

because, if there is some efficacy, then they are more likely to have been infected before being vaccinated than after (relative to a vaccine that has no effect: $\Pr(d_j > D_c | x_j = 0) = q_j$). We solve this iteratively for $\hat{\eta}$ with reference to all observations j.[8] Then

$$\omega_j = \begin{cases} 1 & y_j = 0 \\ \Pr(d_j > D_c | x_j) & y_j = 1 \end{cases} \tag{6}$$

The resulting weights are used in Equations 1, 2, 3, and 4 as before.

Determining the inclusion weight for a vaccinated person

The inclusion weight for a vaccinated person j ($q_j = 1$), whose symptoms began after D_c (randomisation date plus an assumed time from vaccination to seroconversion) is the probability that they were infected after D_c. To determine this, first we make explicit the condition that their infection date d_j is less than their symptom date D_s. The conditional probability we want can then be decomposed into probabilities unconditional on the symptom date, as follows,

$$P(d_j > D_c | q_j = 1, d_j < D_s) = P(D_c < d_j < D_s, q_j = 1) / P(d_j < D_s, q_j = 1)$$

Then splitting the denominator into the probabilities of being infected in two different periods (before versus after D_c) gives

$$P(d_j > D_c | q_j = 1, d_j < D_s) = \frac{P(D_c < d_j < D_s | q_j = 1)}{P(d_j < D_c | q_j = 1) + P(D_c < d_j < D_s | q_j = 1)}$$

Since the probability of being infected before D_c doesn't depend on whether person j was vaccinated, $P(d_j < D_c | q_j = 1) = P(d_j < D_c | q_j = 0)$. We additionally assume that

$$P(D_c < d_j < D_s | q_j = 1) = \psi P(D_c < d_j < D_s | q_j = 0)$$

where ψ is the relative risk of infection between a vaccinated and unvaccinated person, assumed to be constant through time, giving

$$P(d_j > D_c | q_j = 1, d_j < D_s) = \frac{\psi P(D_c < d_j < D_s | q_j = 0)}{P(d_j < D_c | q_j = 0) + \psi P(D_c < d_j < D_s | q_j = 0)}$$

Expressing the right hand side in terms of probabilities conditional on symptoms, by dividing the numerator and denominator by $P(d_j < D_s | q_j = 0)$, then gives the weight for a vaccinated person j as

$$\omega_j = P(d_j > D_c | q_j = 1, d_j < D_s) = \frac{\psi \omega_j^{(0)}}{(1 - \omega_j^{(0)}) + \psi \omega_j^{(0)}}$$

where $\omega_j^{(0)} = P(D_c < d_j < D_s | d_j < D_s, q_j = 0) = P(D_c < d_j < D_s | q_j = 0) / P(d_j < D_s | q_j = 0)$ is the weight if person j were unvaccinated.

[8] This is solved as in expectation maximisation (EM). EM algorithms have been used and discussed in clinical trials for e.g. sample-size re-estimation (Gould and Shih, 1992; Friede and Kieser, 2002; Teel et al., 2015; Huang et al., 2018).

C Supplementary material

Table 10: Analysis methods for trials simulated under the "binary" method framework corresponding to Table 2. We report the total weights for vaccinated and control confirmed cases. The row "True at recruitment" is the result we would obtain with perfect knowledge of who was infected on the day of recruitment. "True" is the result we would obtain if we had perfect knowledge of the day of infection and the day of seroconversion.

Method	Power	Type 1 error	Vaccinated	Control	VE estimate
None	0.59	0.07	16.5	27.8	0.39 (0.19)
Binary	0.74	0.04	6.9	18.0	0.59 (0.19)
Continuous (without VE)	0.75	0.03	6.6	17.0	0.59 (0.18)
Continuous (with VE)	0.81	0.04	5.7	17.0	0.64 (0.19)
True at recruitment	0.72	0.04	8.5	19.7	0.55 (0.18)
True	0.93	0.09	5.2	19.7	0.72 (0.14)

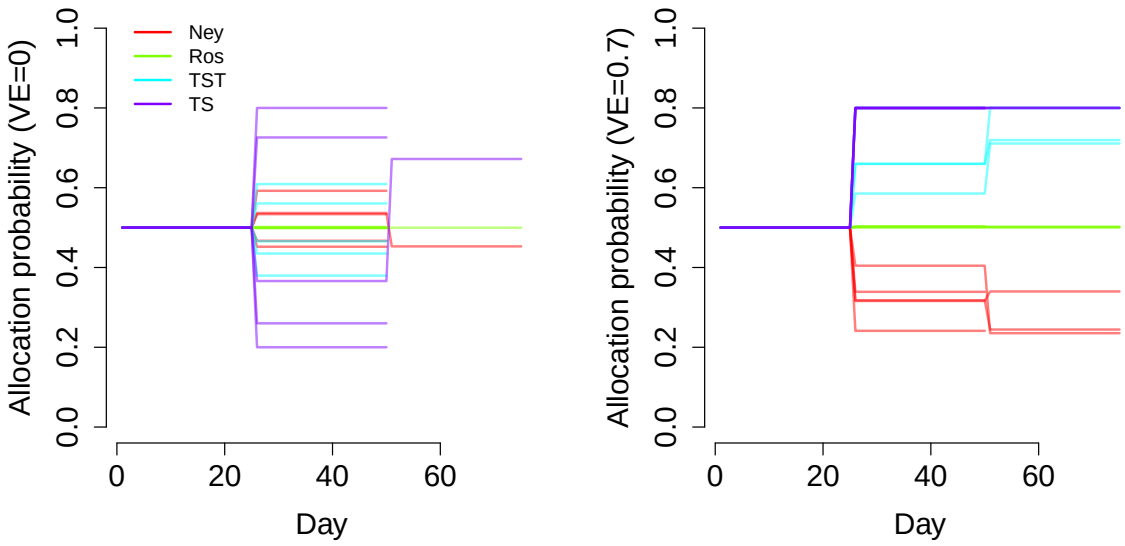

Figure 4: Five samples of trajectories of the allocation probability over time for the adaptive designs.

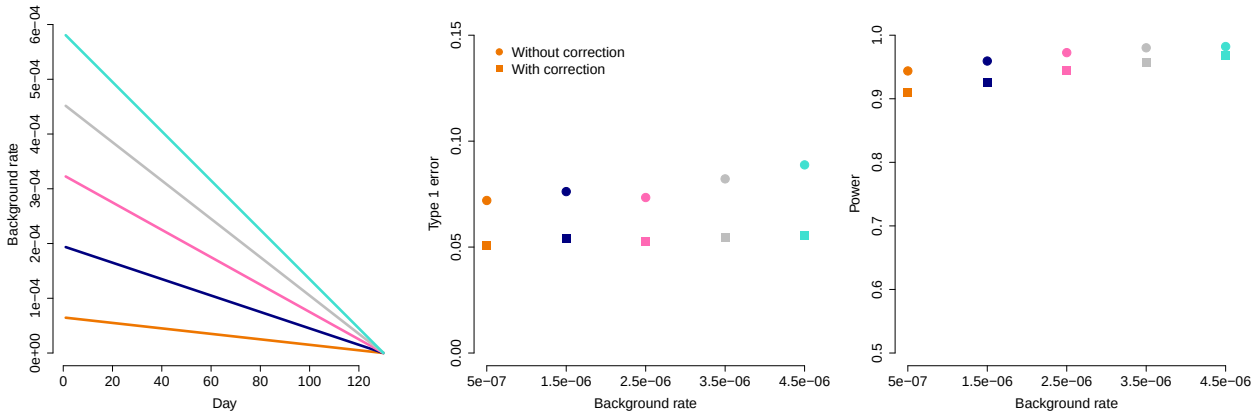

Figure 5: The robustness of one adaptive trial design (TST) to a time trend, demonstrating the effect of the correction of Simon and Simon (2011). Type 1 error rate and power as a function of the trend in the background rate for a trial with a response-adaptive randomisation rate. "Background rate" can be interpreted as the rate of infection by individuals unknown in the context of the trial. (E.g. source population or unknown contact.) We illustrate with linear trends rather than something more realistic, undulating, or stochastic in order to stress test the method. In addition, we choose a trend that we would expect to most favour the Thompson sampling methods. As TS is a more "aggressive" response-adaptive method than TST, we expect it to be more susceptible to time trends, and therefore the loss in power following correction to be exaggerated. Left: Five different time trends for the background rate. The trend is that background rates diminish over time to zero. Middle: The gradient of the trend is shown on the x axis. On the y axis is the type 1 error rate. Right: The gradient of the trend is shown on the x axis. On the y axis is the power.

D Glossary by letter

Letter	Meaning	Letter	Meaning	Letter	Meaning	Letter	Meaning
α	Significance threshold	a	node attributes	A			
β	per-contact infection or transmission rate	b		B			
γ	infectious period	c		C	control-arm label		
δ	recruitment time	d	incubation time	D			
ϵ	enrolment rate	e	trial end	E	exposed	$\mathcal{E}$	set of edges
ζ		f	fails	F			
η	vaccine efficacy	g		G		$\mathcal{G}$	graph
θ		h	"household contact" label	H		$\mathcal{H}$	household
ι		i	index over node	I	infectious		
		j	index over node	J			
κ		k	instantaneous rate of infection	K			
λ		l	Truncated normal distribution parameter	L		$\mathcal{L}$	number infectious contacts
μ		m		M		$\mathcal{M}$	number symptomatic contacts
ν		n	"unpredictable contact" label	N	total	$\mathcal{N}$	neighbourhood
ξ	incubation period	o		O		$\mathcal{O}$	size
π	allocation probability	p	efficacy probability	P			
		q		Q			
ρ	allocation ratio	r		R	removed		
σ		s	day of symptoms	S	susceptible		
τ	seroconversion time	t	exposure time	T			
υ		u		U	unenrolled label		
ϕ	Thompson tuning parameter	v	trial arm	V	experimental-arm label	$\mathcal{V}$	set of nodes
		w	"workplace contact" label	W		$\mathcal{W}$	workplace
χ	edge weight	x	vaccination label	X			
ψ	relative risk of infection given vaccination	y	confirmed-case label	Y			
ω	inclusion weight	z		Z			