Limitations of Disease X vaccine efficacy and safety clinical trials

Julien Flaig* Nicolas Houy†

October 11, 2025

Abstract

Background: In a vaccine clinical trial, the candidate is typically accepted or rejected based on predefined efficacy and safety thresholds. An efficacy threshold is the minimum measured treatment effect deemed statistically significant and clinically relevant. A safety threshold is the maximum number of adverse events deemed acceptable in the vaccine arm of the trial. However, the uncertainties and changing conditions met during the emergence of an unknown infectious disease (Disease X) may hinder such simple approaches.

Method: We model the emergence of a Disease X with a SIR (susceptible-infectious-recovered) transmission model. A vaccine is available and the objective is to minimize the total cost over the course of the epidemic, which includes infection, vaccination and adverse event costs. Uncertainties regarding transmission, vaccine, and cost parameters are represented by prior distributions. We simulate placebo controlled efficacy and safety trials, and different vaccination policies. Under policies using trial results, the susceptible population is vaccinated if and only if the vaccine candidate passes both the efficacy and safety tests.

Results: In our baseline scenario and on average over uncertain parameters, using clinical trial outcomes to decide whether to vaccinate the population yields a lower expected total cost compared to indiscriminate emergency vaccination and to no intervention. However, testing both efficacy and safety yields a higher cost compared to testing safety alone. The difference counts in billions of dollar for a population of 10^8 individuals. In our scenario, there is an optimal intermediate safety threshold value that better discriminates vaccine candidates. By contrast there is no optimal intermediate efficacy threshold value. The incidence difference between the treated and control arms is even a misleading measure of vaccine performance.

Conclusion: We show with an example that simple threshold-based decision rules may not be appropriate to discriminate candidates in a Disease X vaccine clinical

*Spegegant, F-69002, Lyon, France. Email: corr.author.flaig@spegegant.com.

†University of Lyon, Lyon, F-69007, France; CNRS, GATE Lyon Saint-Etienne, F-69130, France. Email: houy@gate.cnrs.fr

trial. Uncertainties and the disease dynamics need to be fully taken into account, which may require more advanced pattern recognition methods.

Keywords: Disease X, emerging disease, vaccination, public health, vaccine efficacy, efficacy safety, clinical trial simulation, uncertainty.

1 Introduction

Disease X is a hypothetical yet unknown disease with epidemic potential. It has been included in WHO's list of priority diseases for research and development since 2018 [1, 2]. Responding in a timely manner to a Disease X outbreak is challenging due to the potentially fast dynamics of the epidemic and pervasive uncertainties – whether about the pathogen and its transmission, or about prospective interventions. This also applies to emerging and reemerging diseases. Vaccine development in such scenarios has received notable attention following the 2013-2016 Ebola crisis [3, 4], and since then with the development and approval of COVID-19 vaccines at an unprecedented speed [5] and a continued aim at further acceleration [6, 7].

Vaccine candidates are approved for use based on the assessment of their overall benefits and risks in target populations [8–10]. This assessment typically consists of a series of steps from preclinical studies to the clinical stage. The final step prior to approval decision is a phase III randomized controlled trial (RCT) whereby the candidate is administered to a large group of volunteers, the treatment group, while another group, the control group, receives a placebo or a comparator vaccine. The candidate's efficacy is assessed by comparing disease incidence in the treatment group with disease incidence in the control group [11]. Vaccine safety is monitored by comparing the incidence of adverse events of interest in the control group and in the general population [12, 13]. While ever more advanced methods are being developed for the design and result analysis of RCTs [14], simple efficacy and safety estimands are still used overwhelmingly (see for instance published FDA approval data [15]).

In this article, we use numerical simulations to show with an example that a phase III-like clinical trial may only poorly inform approval and public health decisions in a Disease X outbreak scenario. We model the spread of the disease with a susceptible-infectious-recovered (SIR) model – a minimal model structure widely used in textbooks to illustrate epidemiological concepts – with parameter uncertainty represented by probability distributions. In our simulated scenarios, a vaccine approval decision is made based on efficacy and safety hypothesis testing in a placebo controlled clinical trial. If the vaccine is approved, it is then delivered to the whole susceptible population. This scenario is simulated for a number of randomly drawn uncertain parameter values, so uncertainty is propagated from, say, biological parameters of Disease X down to the public health impact of vaccine approval decision. The objective is to minimize the total cost of the outbreak, which includes infection costs, vaccination costs, and vaccine adverse events costs. Notice that we take a payer's or societal perspective rather than an individual's perspective, which

is an important distinction to make when interpreting our results.

Our illustrative example relies on a specific and stylized emergency response scenario where large-scale vaccine rollout is an option. Still, the underlying reasoning applies to any instance of binary (go/no-go) decision under uncertainty during vaccine development. Issues similar to those pointed here can be expected e.g. for decisions based on correlates of protection [16] or non-clinical endpoints (biomarkers) that could be considered in an emergency response scenario. Our scenario is optimistic in this respect, as we assume that a clinical trial is feasible under favorable conditions, which may of course not always be the case. Under this assumption, we elaborate on an example where clinical trial efficacy and safety estimates are actually misleading on average over uncertain parameters. That is, they do not provide "minimum information" [17] to be improved upon, e.g. through post-approval surveillance or the collection of real world evidence (RWE, see e.g. [18]), but may rather lead to poor and potentially irreversible public health decisions. The claim that clinical trials are imperfect yet better than nothing does not hold in our scenario. Similarly, arguments in favor of efficacy as "an intrinsic property of the vaccine (...) that can be generalized from one setting to another" [19] fall, as our results suggest potential poor performance of approval decisions even assuming that RCT and rollout are carried out in the same setting (but with dynamic effects in the way). We argue that the issue is not specific to our particular example of clinical trial data (mis)use, and goes beyond incremental methodological or technical fixing.

The design and use of clinical trials to support decision making has previously been addressed extensively and across therapeutic areas. General articles (for instance [20] or more recently [21]) provide lists of design mistakes to be avoided, warnings against RCT data misinterpretation, or RCT limitations to overcome. These issues can be solved technically, and the articles' focus lies accordingly on the technical difficulty of properly estimating efficacy and safety. This misses the point of the actual value of efficacy and safety estimates to support decision making. As for more specific quantitative studies, they are most often solution-oriented and focus each on a limited number of design or analysis issues in specific scenarios. By contrast, we use a quantitative minimal example to make a general point about the value and use of RCT data for decision making. We control merely technical issues by assuming favorable trial conditions, which allows us to dive into more fundamental problems.

Decision making, in our scenario, is undermined by a variety of interrelated factors. One conspicuous barrier to using clinical trial data for proper decision making is that the eventual net benefit of a vaccination campaign depends on parameters that may not be easily (and are typically not) assessed in a clinical trial. This is usually the case of

economic parameters in a broad sense (e.g. logistical costs or constraints, averted loss of productivity), but also societal benefits [17, 22, 23]. Collecting economic data alongside otherwise typical clinical trials ("piggybacking") has long been discussed [24–28] but is not yet widespread [29]. At any rate, it is reasonable to assume remaining uncertainties about economic parameters at the time of approval decision in a Disease X outbreak scenario.

This uncertainty is compounded by uncertainties regarding parameters that *do* influence clinical trial results, for example pathogen, vaccine, or population characteristics, including the potential for indirect vaccine protection, which is typically ignored in vaccine trials [17, 22, 30]. The total net benefit of a vaccination campaign depends on the long-term trajectories of the epidemic with and without vaccination (that is, the difference of the two), that are influenced by these uncertain parameters. Yet parameter values yielding very different dynamics in the long run may result in similar clinical trial results. This issue can be more serious in a Disease X outbreak scenario, because the early stages of an epidemic do not reflect plainly its long-term dynamics in general [31–33].

We emphasize the distinction between two sources of uncertainty that we both tackle with Monte Carlo simulations in our study: first, parameter uncertainty, that is the fact that the true value of the parameters is not known, and, second, uncertainty due to trial random sampling e.g. of a heterogeneous population or of a stochastic process in an otherwise homogeneous population.[1] Traditional clinical trial design methods have mostly ignored the former and focused on the latter (despite recommendations, see e.g. [26]). Take for example textbook sample size calculations: given a working significance level, and assuming a treatment efficacy value (e.g. the smallest worthwhile value, which is arguably an uncertain parameter at the time of the trial), a sample size is picked to achieve a target power – the limitations of this "conditional power" approach have been discussed by other authors, e.g. [31, 34].[2] The external validity of trials has largely been equated with proper sampling (inclusion and sample size) of a possibly heterogeneous population e.g. in scholarly discussions about *efficacy* (performance under trial conditions) versus *effectiveness* (performance in the field) of treatments [20]. A large portion of the clinical trial design literature focuses on tweaking existing methods so that sample efficacy and safety estimates accurately reflect efficacy and safety in the real world (that is, improve sampling to avoid biases). However, the actual value of more accurate efficacy and safety estimates to inform decisions is rarely considered in practice. To be fair, an important body of earlier

[1] Notice that this distinction is somewhat artificial. Think of the potential influence of the clinical trial (vaccinated individuals in the treatment arm) on the course of epidemic: it is an uncertain parameter of transmission that depends on trial random sampling.

[2] But we note that it still seems to have influenced simulation-based vaccine trial design studies that typically assume an efficacy value, or consider a limited number of efficacy scenarios [35–42].

studies used *value of information* reasoning to determine optimal sample sizes considering the cost and benefit of sampling, but in static settings and using analytical methods that would be inappropriate in an infectious disease scenario [43–52] – we refer to [33] for an illustration of the complex dynamics of value of information in a Disease X emergence scenario. In the present study, we go beyond the usual efficacy versus effectiveness distinction and the corresponding problem of appropriate sampling. Specifically, we assume a large sample size in our base case, and minimal population heterogeneity. We make the optimistic assumption that the sampled and general populations have the exact same characteristics, save for vaccination date, and that the only inter-individual differences are differences in stochastic trajectories (infection, recovery) and treatment (vaccination). However, all model parameters are uncertain.

The introduction of Bayesian methods in clinical trials has been a significant step towards proper consideration of parameter uncertainties. Bayesian methods offer ways to aggregate information as it accumulates during a trial [53, 54]. Prior distributions over uncertain parameters are updated into posterior distributions that can in turn be sampled to extrapolate from available information and support decision making e.g. through cost-effectiveness analyses [55, 56]. More generally, Bayesian methods allow to "move from the 0–1 utilities to utilities that put values on units of the outcome(s) of interest" (Stangl [57] in her discussion of the seminal paper by Brophy [58]). Implicit 0–1 utilities are found in frequentist hypothesis testing where the hypothesis of treatment effect is accepted or rejected depending on whether the observed difference between the treatment and control groups lies on one side or the other of a predefined threshold. Examples of threshold-based criteria include requirements for COVID-19 vaccine issued in 2020 [59–61]. Thresholds are known to lose information [49, 62, 63]. This loss of information is twofold: first, hypotheses concern only a limited number of summary estimands, while the decision problem involves many parameters as discussed above[3] and second, information is lost about the level of parameter uncertainty.

Still, Bayesian methods are not widely used in clinical trials [54], and not always to their full potential. For example, Bayesian clinical trials often consist in producing the posterior distribution of a single summary estimand (e.g. efficacy), from which extrapolation can be difficult – see [65] for instance. Similarly, the value of information studies cited above [43–52] did build upon Bayesian reasoning, but with a limited number of uncertain parameters and a limited propagation of uncertainty. As for implications, they mostly criticized thresholds used in hypothesis testing for being arbitrary; we go a bit further and

[3]This issue remains the same irrespective of the considered summary estimand. For example, *effectiveness* estimands as presented e.g. in [11, 64] would likely have similar limitations. For efficacy testing, the present article looks primarily into incidence difference between RCT arms without loss of generality.

show that they may also be misleading. Finally, it is worth mentioning the substantial Bayesian literature on predictive probabilities of trial success (including frequentist trials) to support interim decision making [34, 53, 66, 67]. Interestingly, some authors in this literature propagated biological uncertainties down to trial performance and considered a larger number of parameters, but not economic parameters [68, 69]. While we take a public health perspective in this article, some of our comments may still apply to trial sponsor decision making.

Our article is closer to some previous studies. Hill-McManus and Hughes [63] used RCT simulations to pick a trial design (inclusion criterion, sample size, and medicine doses) maximizing the manufacturer's return on investment. They propagated biological uncertainties down to the marketing stage, and modeled reimbursement decisions based on a cost-effectiveness model. While they considered a wide range of uncertain parameters and did include economic parameters, those are not uncertain in their study. They also assumed seamless integration of RCT results with health economic modeling and reimbursement decision, which may not be the rule in practice. To some extent, our work is in line with Perkins et al. [70] (unpublished to our knowledge). This study argued that "statistical uncertainties" (i.e. around efficacy estimates) and "biological uncertainties" (i.e. about the disease and vaccine mode of action) make it difficult to make public health decisions based on the results of a dengue vaccine clinical trial alone. Our methods differ, however, as they back-calculated uncertain parameter distributions based on the results of a real clinical trial. By contrast, we start from hypothetical uncertain parameter distributions and simulate RCTs *as if* we were faced with Disease X outbreaks. They also focused on vaccine efficacy and ignored safety while we consider both, and did not include economic parameters while we do.

The remaining sections of the article are as follow. Section 2 introduces all modeling assumptions and modeled scenarios. The results are presented in Section 3. Section 3.1 shows how making a decision based on RCT results compares to other policies in our scenario. In Section 3.2, we vary the design of RCTs – Section 3.2.1 is the core of the article and looks into statistical tests, while Section 3.2.2 describes more concisely the influence of trial duration and sample size. Section 4 concludes.

2 Materials and methods

2.1 Disease X transmission model

We consider the emergence of an unknown infectious disease, Disease X, in a closed homogeneous population of $N = 10^8$ individuals. We assume that the spread of the disease

can be described with a deterministic SIR (susceptible-infectious-recovered) compartmental model with basic reproduction model R_0 and $1/\gamma$ average duration of infectious period. Figure App-1 shows a sketch of the transmission model.

At date 0, a single individual is infected and all other individuals are susceptible. A vaccine is available from date $T_v = 900$ days (unless otherwise specified). We assume no alternative vaccine or treatment. The vaccine is administered as a single dose and protects vaccinated individuals with probability p. Vaccine protection is immediate and there is no waning of protection. Adverse events occur in vaccinated individuals with probability ϵ, in which case the average duration between vaccination and adverse events is δ. Adverse event probability ϵ is independent of vaccination success.

The cost of infection is c_i per infectious day, and the cost of adverse events is c_{ae} per case. The total cost of producing and administering one vaccine dose is c_v – we ignore fixed costs for concision, which amounts to assuming that they have already been incurred at time T_v.

2.2 Decision problem

At time T_v, a decision must be made by the decision maker (health authorities or the payer) between several options. They can implement vaccination for the entire susceptible population immediately (emergency vaccination); they can decide to not vaccinate anyone; or they can implement a RCT first and implement vaccination later conditionally on the RCT outcomes (by ways to be defined below). We assume that the objective of the decision maker is to minimize the expected total cost over the course of the epidemic.

2.3 Uncertainties and prior information

We consider parameter uncertainties at time T_v represented by the parameter prior distributions shown in Table 1. The population size N and the date T_v from which the vaccine is available are assumed to be known. We are only interested in decision making from date T_v. The value of T_v clearly has an influence on the trade-off between emergency vaccination and clinical research, but this was investigated elsewhere [33] and is left outside the scope of the present study. Figure 1 shows 400,000 parameter draws from the prior distributions of uncertain parameters. We will simulate decision making for each parameter draw.

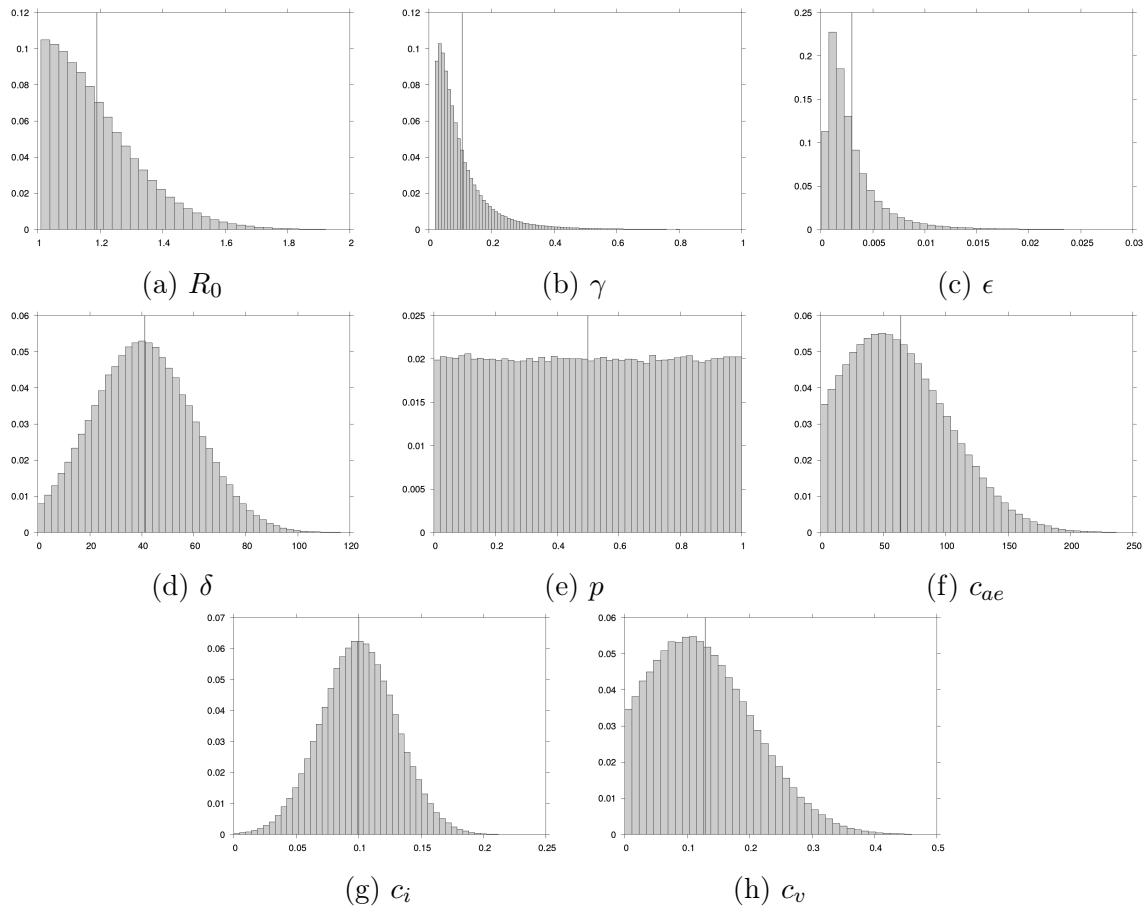
Prior information includes knowledge of the model structure, that is SIR transmission in the absence of intervention, and knowledge of the vaccine characteristics (all-or-nothing protection, no waning). These assumptions can readily be relaxed by extending the model

with additional compartments and uncertain transition rates, or by introducing several models with a probabilistic weight for each. In this didactic article, the SIR structure is to be understood as a minimal working model of transmission in a population.

Parameter	Distribution	Truncation
N	10^8 (fixed)	
T_v	900 (fixed)	
R_0	Gamma distribution with mean 1.05 and standard deviation 0.2	$[1.01; \infty]$
γ	Lognormal distribution with mean 0.1 and standard deviation 0.1	$[0.02; \infty]$
p	Uniform distribution	$[0, 1]$
ϵ	Lognormal distribution with mean 0.003 and standard deviation 0.003	$[0; 1]$
δ	Normal distribution with mean 40 and standard deviation 20	$[0; \infty]$
c_v	Normal distribution with mean 0.1 and standard deviation 0.1	$[0; \infty]$
c_i	Normal distribution with mean 0.1 and standard deviation 0.03	$[0; \infty]$
c_{ae}	Normal distribution with mean 50 and standard deviation 50	$[0; \infty]$

Table 1: Model parameters prior distributions. Time unit: day. Cost unit: 10^3 USD.

We choose biologically plausible disease parameter values (R_0, γ). The values of ϵ and δ are assumed, while the prior distribution of p represents minimal prior information on vaccine efficacy. The order of magnitude of ϵ may seem high (see [71, 72]), but recall that it represents the probability of a "synthetic" adverse event that we hypothesized for concision, whose probability is the sum of several "real" adverse events of interest. The order of magnitude of the cost c_{ae} of adverse events is based on damage payments in the US and the UK, with lower range values representing mild uncompensated adverse events [73, 74]. The order of magnitude of the cost c_i of an infectious day is that of average daily wages in high-income countries [75]. The cost per administered vaccine dose c_v derives from vaccine purchase prices [76, 77]. We emphasize that prior distributions represent available information or beliefs *before* vaccine approval decision. In addition, this information concerns one particular outbreak and one particular vaccine candidate. For these two reasons, prior distributions may not correspond to distributions observed among approved vaccines or inferred from past epidemiological data.



(a) R_0 (b) γ (c) ϵ

(d) δ (e) p (f) c_{ae}

(g) c_i (h) c_v

Figure 1: Frequency distributions for 400,000 random draws of parameter values. Vertical lines show average values. Time unit: day. Cost unit: 10^3 USD.

2.4 Simulated RCTs and trial simulation method

We simulate two-arm clinical trials starting at date T_v with duration T_{ct}. The default trial duration is $T_{ct} = 80$ days. The sample size is N_{ct} (default $N_{ct} = 50{,}000$), and the treatment and control arms have equal size $N_{ct}/2$. Individuals in the treatment arm receive the vaccine, while individuals in the control arm receive a placebo.

The number of infections in the treatment and control arms, n_1^I and n_0^I respectively, and the number of adverse events n_{ae} in the treatment arm are tracked for the duration of the trial T_{ct}. In trial simulations, infections are randomly drawn at each time step in a binomial distribution with probability $R_0 \gamma I/N$ among the remaining susceptible individuals in the considered arm. The number n_{ae} of adverse events in the treatment arm is drawn in a binomial distribution with parameters $N_{ct}/2$ and $\epsilon \left(1 - e^{-T_{ct}/\delta}\right)$.

Vaccine efficacy and safety are then tested at time $T_v + T_{ct}$. We test whether vaccine

efficacy is *superior* [78] to that of the placebo using a z-test, where the null hypothesis $H_0 : n_1^I = n_0^I$ is tested against hypothesis $H_1 : n_1^I < n_0^I$ with significance level α. The rate of occurrence of adverse events is estimated as $\lambda_{ae} = -\frac{1}{T_{ct}} \log \left[1 - \frac{2n_{ae}}{N_{ct}}\right]$. The vaccine passes the safety test if λ_{ae} is below a threshold λ_{ae}^{max} based on the background incidence of medical conditions of interest in the unvaccinated population. See [12, 71, 72, 79, 80] for illustrations and discussions of this approach.

In the following we will focus on two particular clinical trials, denoted CT95 and CT0. In both trials, the sample size and duration are set to their default values $N_{ct} = 50{,}000$ and $T_{ct} = 80$ days. The safety threshold is set to $\lambda_{ae}^{max} = 2.5 \times 10^{-5}$ day^{-1}.[4] CT95 and CT0 differ in the significance level used in efficacy testing. For CT95, significance is set to customary value $\alpha = 0.05$ (95% confidence), while for CT0 it is set to $\alpha = 1$ (0% confidence). With CT0, all vaccines pass the efficacy test and may only be rejected on the basis of safety.

3 Results

3.1 Use of RCT information

The objective of a RCT is to evaluate the safety and efficacy of a vaccine candidate. Table 2 shows how this information can help reduce the total cost over the course of an epidemic. We estimate the expected total costs over prior parameter distributions in different intervention scenarios (Figure App-2 shows the full cumulative distributions of total costs). The simulated intervention scenarios are as follow. In scenario "no trial, no vacc." no one is vaccinated. "No trial, vacc. at T_v" corresponds to emergency vaccination: all susceptible individuals are vaccinated at time T_v i.e. as soon as the vaccine is available. Under policies "CT0" and "CT95", a clinical trial is implemented at time T_v. If the vaccine passes the safety and efficacy tests, as defined in Section 2.4, all susceptible individuals receive the vaccine at time $T_v + T_{ct}$.

In scenarios "trial, no vacc." and "trial, vacc at $T_v + T_{ct}$", a trial is implemented at time T_v but the vaccine is used or not irrespective of the trial result. Comparing "no trial, no vacc." and "trial, no vacc." shows that the influence of the clinical trial (that is, of vaccinating $N_{ct}/2 = 25{,}000$ individuals at time T_v) on disease transmission is small in our scenario. Similarly, comparing emergency vaccination ("no trial, vacc. at T_v") to "trial, vacc. at $T_v + T_{ct}$" shows that delaying vaccination by $T_{ct} = 80$ days to implement a RCT has little influence on the course of the epidemic.

Making a decision based on CT0 or CT95 allows us to reduce the overall cost com-

[4] 912.5 cases per year per 100,000 individuals.

pared to making a decision irrespective of trial results. Compared to "trial, no vacc.", the increase in vaccination and adverse event costs is more than compensated by the decrease in infection costs. Compared to "trial, vacc. at $T_v + T_{ct}$" and to emergency vaccination, the increase in infection costs is more than compensated by the decrease in vaccination and adverse event costs.

"Trial, optimal at $T_v + T_{ct}$" is an hypothetical scenario in which a clinical trial is implemented at time T_v, and the true value of all uncertain parameters is known at time $T_v + T_{ct}$, so that the best policy can be picked between vaccinating and not vaccinating the remaining susceptible individuals. This gives a (loose) lower bound on the total cost achievable through clinical research. Using the no intervention scenario ("no trial, no vacc.") as a baseline, the value brought by CT0 and CT95 is worth about 31 % and 7 % respectively of the value of perfect information.

	Cost			
Policy	**Infections**	**Vaccination**	**Adverse events**	**Total**
No trial, no vacc.	44.25 (44.10 – 44.39)	0 –	0 –	44.25 (44.10 – 44.39)
No trial, vacc. at T_v	16.41 (16.32 – 16.50)	11.10 (11.08 – 11.13)	16.47 (16.40 – 16.54)	43.98 (43.87 – 44.09)
Trial, no vacc.	44.22 (44.07 – 44.36)	3.22×10^{-3} (3.22×10^{-3} – 3.23×10^{-3})	4.08×10^{-3} (4.06×10^{-3} – 4.10×10^{-3})	44.22 (44.08 – 44.37)
Trial, vacc. at $T_v + T_{ct}$	17.78 (17.68 – 17.87)	10.98 (10.95 – 11.00)	16.28 (16.21 – 16.36)	45.04 (44.83 – 45.15)
Trial, optimal at $T_v + T_{ct}$	22.62 (22.52 – 22.71)	3.54 (3.52 – 3.56)	3.72 (3.69 – 3.74)	29.87 (29.78 – 29.97)
CT0	29.76 (29.63 – 29.88)	6.01 (5.99 – 6.04)	3.98 (3.96 – 4.00)	39.75 (39.63 – 39.87)
CT95	40.13 (40.00 – 40.27)	1.84 (1.82 – 1.85)	1.21 (1.20 – 1.22)	43.18 (43.04 – 43.32)

Table 2: Summary simulation results for different vaccination policies. Costs in 10^9 USD. Averages over 400,000 parameter draw, and one simulation per parameter draw. Parentheses: 95% confidence intervals.

Noticeably, we obtain a better result with CT0 compared to CT95. Recall that CT0 only tests safety, while CT95 also tests efficacy with significance level $\alpha = 0.05$. A decision maker would arguably be more likely to implement the customary CT95 trial rather than CT0 *a priori*, that is without performing the analyses presented in Table 2. It is critical to fully consider uncertainties in these analyses. Table App-1 is the equivalent of Table 2 but with uncertain parameters set to their prior expected value. On this basis, the best

policy would be to do nothing.

CT0 and CT95 are compared in more details in Table 3. In our scenario, vaccination is optimal at time $T_v + T_{ct}$ for 31.80 % of uncertain parameter values, but CT0 and CT95 poorly prescribe vaccination when it is actually optimal. The true positive rate of CT95, in particular, is only $6.16/31.80 \approx 0.19$, and that of CT0 is $20.47/31.80 \approx 0.64$. Because CT95 is more restrictive, it has a lower false positive rate of $10.86/68.20 \approx 0.16$ compared to CT0 ($34.25/68.20 \approx 0.50$).

		Vaccination	**No vaccination**	**Total**
CT0	**Vaccination**	20.47 %	34.25 %	54.72 %
	No vaccination	11.33 %	33.96 %	45.29 %
	Total	31.80 %	68.20 %	100.00 %
CT95	**Vaccination**	6.16 %	10.86 %	17.02 %
	No vaccination	25.63 %	57.34 %	82.97 %
	Total	31.80 %	68.20 %	100.00 %

Table 3: Correlation matrix between vaccination policy implemented based on CT0 and CT95 (row) and vaccination policy implemented under perfect information after a clinical trial (column). Percentages of 400,000 parameter draws.

3.2 Clinical trial design

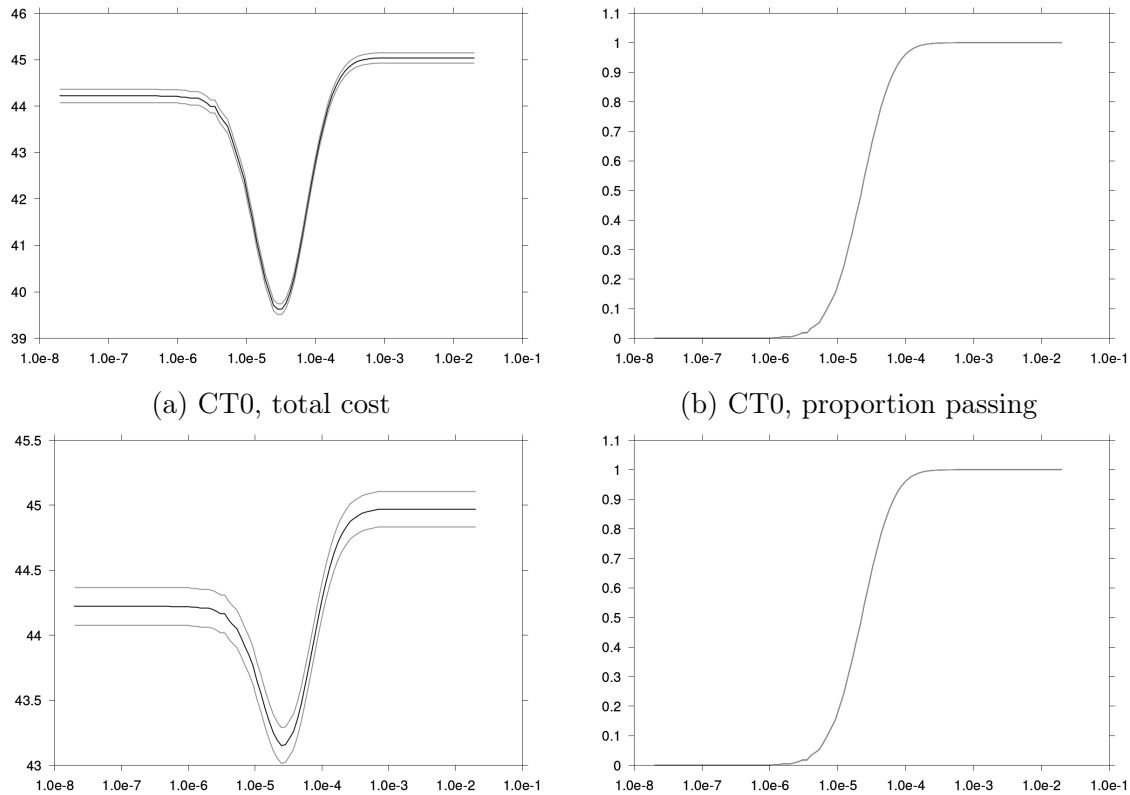
The performance of a clinical trial depends on its design. In this section we ask whether the performance of CT0 and CT95 could be improved on average over uncertain parameters by changing the statistical tests, the sample size, or the trial duration.

3.2.1 Statistical tests

Here, we first describe how safety and efficacy threshold values used in statistical tests may increase or decrease the expected total cost of making a decision on the basis of a RCT, which will tell us about RCT design optimization on average over uncertain parameters. This is shown in Figures 2–4. Then, we describe the relationship between safety and efficacy estimates (i.e. RCT observations), and the total benefit of vaccinating the whole population, *given* uncertain parameter values. This will help understand the previous results and give insights into the level of information provided by a RCT in our scenario. This is done in Figures 5 and 6.

Let us look into the influence of safety threshold λ_{ae}^{max} on clinical trial performance. In Figure 2, we show for both CT0 and CT95 the expected total cost and the proportion of simulations (one simulation for each of the 400,000 parameter draws) in which the

vaccine passes the safety test as a function of λ_{ae}^{max}. If the threshold is too low, the vaccine is rejected with probability 1 and we obtain the same total cost as with policy "trial, no vaccination" in Table 2. If the threshold is too high, the vaccine is rejected with probability 0 on the basis of safety (with CT95, however, it may be rejected on the basis of efficacy) and we obtain the same cost as with policy "trial, vaccination at $T_v + T_{ct}$". The threshold is optimal for an intermediate value.



(a) CT0, total cost

(b) CT0, proportion passing

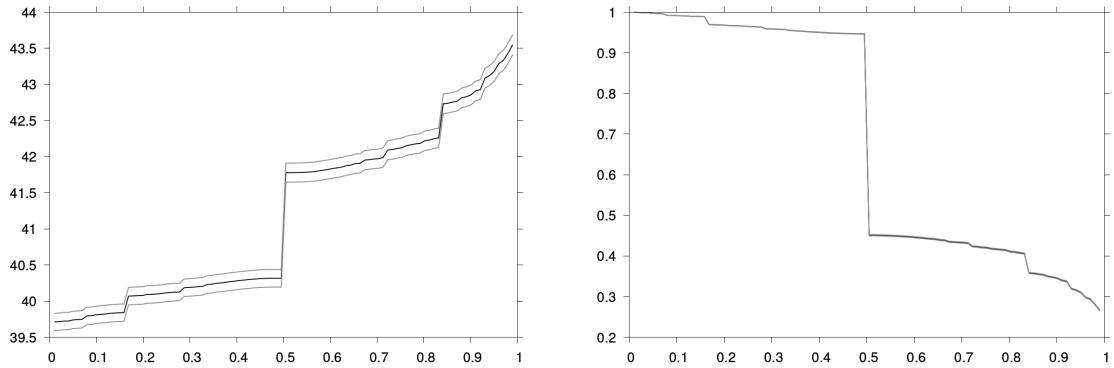
(c) CT95, total cost

(d) CT95, proportion passing

Figure 2: Expected total cost (left panels, 95% CI in light lines, cost in 10^9 USD) and proportion of simulations passing the safety test (right panels) as functions of safety threshold λ_{ae}^{max} ($\log_{10}$ scale).

Let us now turn to the influence of the confidence level $1 - \alpha$ used for efficacy testing. Figure 3 shows the expected total cost and probability of passing the efficacy test as functions of $1 - \alpha$ for our baseline λ_{ae}^{max} value. Recall that $1 - \alpha$ is equal to 0 and 0.95 for CT0 and CT95 respectively. In our scenario, the more restrictive the efficacy test, the higher the overall cost. The curves are not smooth because several simulations may have the exact same difference in number of infections between the treatment and control arms. In our scenario, this is obtained for small numbers of infections in both arms. For example,

there is a single infection case in the control arm and none in the treated arm for more than 4% of simulations, and we obtain the opposite result for more than 2% of simulations. The jump at $1 - \alpha = 0.5$ is explained by a large number of simulations with the same number of cases in both arms. In particular, there is no infection case in either arm for about 32% of the simulations.

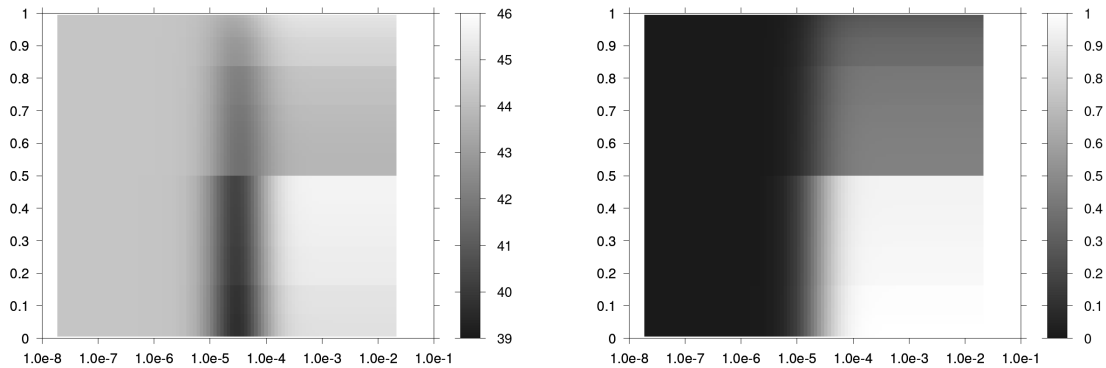


(a) Total cost (b) Proportion passing

Figure 3: Expected total cost (left, 95% CI in light lines, cost in 10^9 USD) and proportions of simulations passing the efficacy test (right) as functions of the efficacy test confidence level $1 - \alpha$.

Figure 4 shows how varying both λ_{ae}^{max} and $1 - \alpha$ affects the results – see also receiver operating characteristic (ROC) curves in Section C in Appendix. The higher expected total cost is incurred for too permissive trials (bottom right quadrant). Starting from there, it is better to optimize λ_{ae}^{max} *alone* rather than increasing stringency of both efficacy and safety testing. For permissive safety tests (right part of the heat map), the trial can be improved by making the efficacy test more restrictive (top right quadrant). However, for intermediate (and optimal) safety threshold values, the more restrictive the efficacy test, the higher the overall cost.

In fact, threshold-based efficacy testing is not a good tool to inform public health policies. Let us dive into how RCT (observed) safety and efficacy estimates relate to the actual optimal policy between vaccinating and not vaccinating after the trial. For 400,000 parameter draws and trial simulations we computed (i) the optimal policy between "trial, no vaccination" and "trial, vaccination at $T_v + T_{ct}$", and (ii) the trial test statistics used for safety and efficacy testing, λ_{ae} and $n_0^I - n_1^I$ (and the efficacy z-value, to be specific) respectively. Figure 5 shows which of the two policies is optimal for each simulation in the safety-efficacy plane. The test thresholds of CT95 are shown in green. In the bottom right quadrant scenarios, the vaccine would be approved on the basis of CT95. In the two top quadrants, the vaccine fails the safety test. In the two left quadrants,



(a) Total cost

(b) Proportion passing both tests

Figure 4: Expected total cost (left) and proportion of simulations passing both safety and efficacy tests (right) as functions of safety threshold λ_{ae}^{max} (x-axis, $\log_{10}$ scale) and efficacy test confidence level $1 - \alpha$ (y-axis).

the vaccine fails the efficacy test. Red dots in the bottom right quadrant correspond to false negatives (the vaccine is rejected while it should be approved). Blue dots in the other quadrants correspond to false positives (the vaccine is approved while it should be rejected). We emphasize that each dot corresponds to a random draw of uncertain parameters and trial sample. In particular, the actual RCT outcomes would correspond to a single dot. Arguably, no pair of threshold values can adequately separate red dots from blue dots in the safety-efficacy plane. This illustrate the limitations of low-dimensional methods when solving a complex high-dimensional problem. The issue is fundamental, and beyond tweaking threshold values in statistical tests.

Now, how do RCTs perform on average depending on threshold values? For the same 400,000 simulations, we computed the difference in total cost between the two policies – "trial, no vaccination" and "trial, vaccination at $T_v + T_{ct}$" –, that is the benefit of vaccinating following the trial. We regressed the benefit of vaccination on test statistics as shown in Figure 6. The regression on λ_{ae} (Figure 6a) crosses $y = 0$ for a cutoff value corresponding to the optimal λ_{ae}^{max} previously highlighted in Figure 2. For a given threshold λ_{ae}^{max}, dots such that $\lambda_{ae} > \lambda_{ae}^{max}$ fail the safety test, and the others pass it. The optimal λ_{ae}^{max} only rejects λ_{ae} values for which vaccination is not optimal on average. The regression on $n_0^I - n_1^I$, by contrast, is strictly negative so there is no intermediate optimal threshold value. Notice also that the regression is strictly decreasing: on average, vaccinating at time $T_v + T_{ct}$ is less beneficial in cases were the reduction of infections in the treatment arm compared to the control arm is *larger*. Hence, as previously shown, the stricter the efficacy test (larger threshold), the worse the average performance of vaccination with vaccines passing the efficacy test (those for which $n_0^I - n_1^I$ is above the threshold).

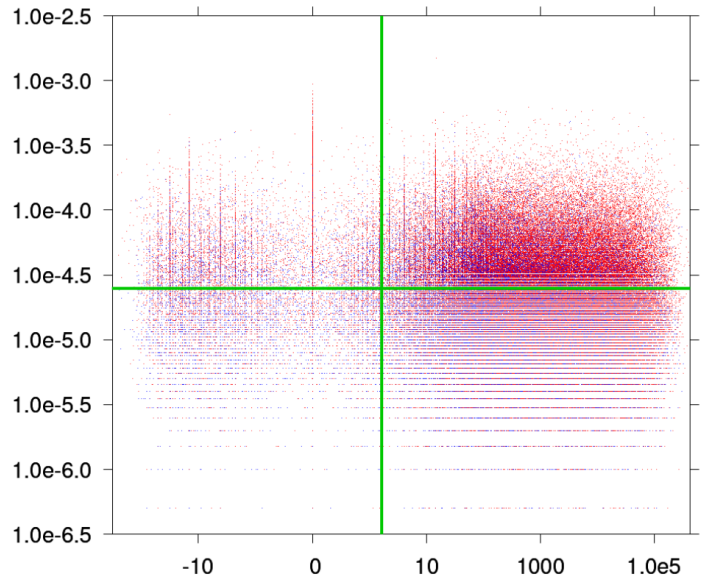
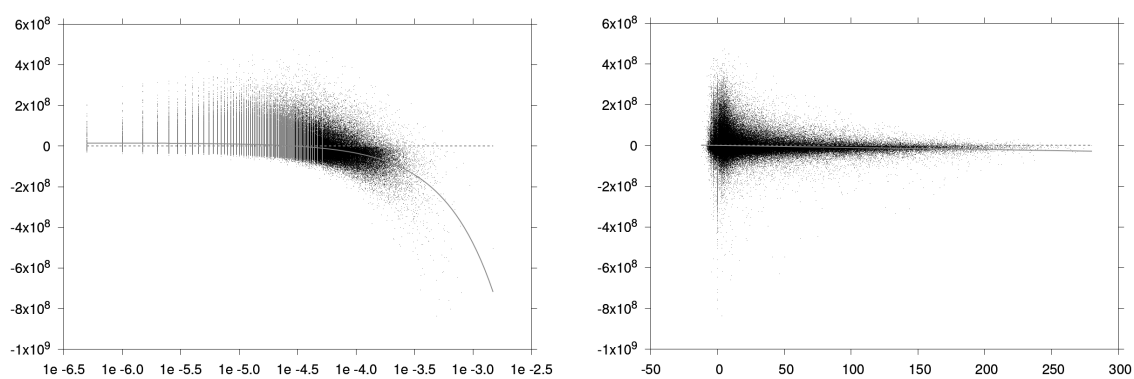


Figure 5: Optimal policy between "trial, no vaccination" (red) and "trial, vaccination at $T_v + T_{ct}$" (blue) as a function of the measured adverse event rate λ_{ae} (y-axis, $\log_{10}$ scale) and the efficacy $(n_0^I - n_1^I)$ z-value (x-axis, symlog scale). 400,000 parameter draws and one trial simulation per draw. Green lines: safety and efficacy thresholds of CT95.



(a) Safety statistic λ_{ae} ($\log_{10}$ scale)

(b) Efficacy statistic $n_0^I - n_1^I$

Figure 6: Difference in total cost over the epidemic between policies "trial, no vaccination" and "trial, vaccination at $T_v + T_{ct}$" as a function of the measured adverse event rate λ_{ae} (left) and the difference in numbers of infections between the control and treatment arm $n_0^I - n_1^I$ (right). 400,000 parameter draws and one trial simulation per draw. Grey line: best fit linear regression. Dashed line: $y = 0$ line.

Figure 6 also illustrates the role of parameter uncertainty. As a thought experiment, assume that all parameters are known and the only source of noise is trial participant

sampling.[5] Assume also that the trial has no influence on the spread of the disease. These assumptions are customary in RCT design, particularly when using a conditional power approach for sample size computation. However, under these assumptions, the scatter plots in Figure 6 would reduce to horizontally aligned dots, that is a cross-section of the full scatter plots. Optimizing the statistical power based only on this limited cross-section can obviously lead to suboptimal trial designs under parameter uncertainty.

3.2.2 Trial duration and sample size

In Figure 7, we show how modifying the duration T_{ct} of the clinical trial or the number N_{ct} of individuals included in the trial affects the results of policies "CT0" and "CT95" consisting in implementing trials CT0 and CT95 respectively at time T_v and vaccinating at time $T_v + T_{ct}$ if the vaccine passed the safety and efficacy tests. Recall our baseline values for CT0 and CT95: $T_{ct} = 80$ days and $N_{ct} = 50{,}000$ individuals.

Many effects come into play when changing the trial duration and sample size. For instance, both parameters influence naturally the quality of information brought by the trial in terms of efficacy as well as safety. But they also influence the spread of the epidemic and the number of individuals still susceptible and hence who can be vaccinated if the vaccine is approved. These individuals are also those who can still suffer adverse events. The resulting total effect on trial performance is difficult to anticipate *a priori*, which highlights the importance of epidemiological simulations in Disease X clinical trial design.

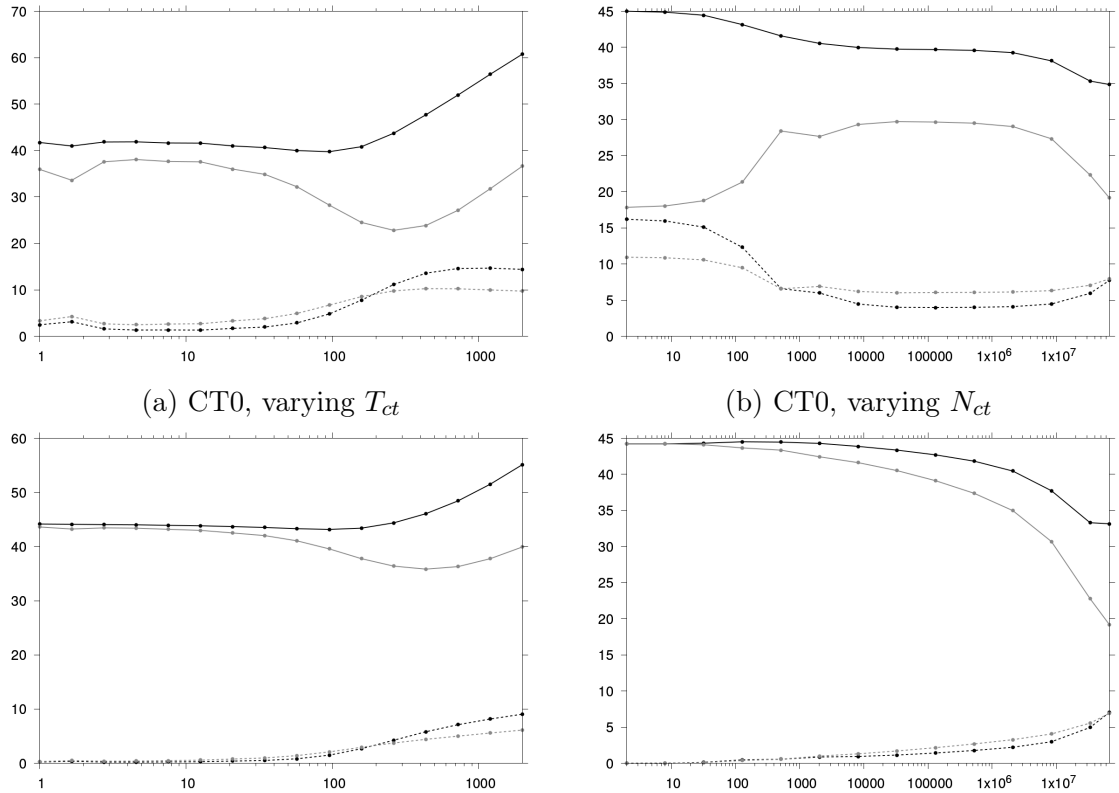
Also, Figure 7 shows that the our baseline T_{ct} and N_{ct} values are such that policies "CT0" and "CT95" perform as good as possible in a realistic range.[6] Hence, the results presented in this article and in particular the poor information given by the RCTs cannot be attributed to poor RCT design.

4 Conclusion

We showed with a numerical example that customary vaccine RCTs with hypothesis testing of efficacy and safety may only poorly inform public health decisions in a Disease X outbreak scenario. We modeled the spread of a hypothetical disease under parameter uncertainty, and we simulated various RCTs and vaccine approval policies. Our analysis considers issues (dynamic effects, uncertainties, large number of parameters involved) that

[5] Recall that we are using a deterministic transmission model.

[6] It seems hard to consider that the entire population can be enrolled in a RCT.



(a) CT0, varying T_{ct}

(b) CT0, varying N_{ct}

(c) CT95, varying T_{ct}

(d) CT95, varying N_{ct}

Figure 7: Expected costs (10^9 USD) as functions of the duration T_{ct} of the clinical trial (days, left) and the number N_{ct} of individuals included in the trial (right) for policies "CT0" (top) and "CT95" (bottom). Black: total cost, grey: infection cost, black dashed: adverse event cost, grey dashed: vaccination cost.

are generally discussed separately, either in general terms, or quantitatively but in a limited setting (e.g. with the objective of solving a specific case). We took a public health perspective, with the objective of minimizing the total cost of the epidemic (including the cost of the RCT). We propagated parameter uncertainties e.g. from biological uncertainties down to the total cost of the epidemic. We simulated counterfactuals that never get to be observed in the field – the hypothetical performance of rejected vaccines. This allowed us to properly estimate the expected total cost of the epidemic under different approval scenarios.

Our analysis relies on several optimistic assumptions – that a RCT with a large sample size is feasible, which may not be the case in a crisis, that population heterogeneity is minimal, and that the RCT and vaccine rollout are carried out in similar settings. These assumptions allow us to overlook technical issues that have already largely been tackled

elsewhere, and to focus with a minimal example on more fundamental issues that may derail decision making.

We simulated two baseline RCTs: a customary safety and efficacy trial, CT95, and a trial testing safety only, CT0. In our scenario, CT0 allowed better decisions than CT95, on average over uncertain parameters. Compared to doing nothing (no RCT, no vaccination), the reduction in total cost with CT0 and CT95 corresponded to about 31% and 7% respectively of the reduction allowed by hypothetical optimal decisions. Testing safety only was a better option on average than testing both efficacy and safety.

In our scenario, safety testing could be optimized to minimize the expected total cost over uncertain parameters. By contrast, and somewhat counter-intuitively, the more stringent the efficacy test, the higher the expected cost of making a decision based on the trial. We looked into how summary efficacy and safety estimates relate to the actual benefit of vaccinating the population for a sample of randomly drawn parameter values. This highlighted the inadequacy of RCT and threshold-based rules to support decisions in our scenario. Finally, we described the complex interaction of different competing effects when varying the trial duration and sample size.

All results presented in this article depend on modeling assumptions, and on the assumed parameter prior distributions in particular. The article is intended to be didactic and should be understood as a call for the development of more advanced pattern recognition methods in RCTs. Put simply, it illustrates that a complex high-dimensional problem may not easily be solved with simple methods leveraging only two dimensions (efficacy and safety).

We believe that our work may have implications beyond Disease X preparedness and outbreak response. Decision making under uncertainty, in a dynamic setting, and with many parameters involved is ubiquitous at all stages of vaccine development. This is not only true in the case of Disease X response. Uncertainty and dynamic effects may arise from the emergence of variants and seasonality, as is the case for the flu. Finally, workarounds such as non-clinical endpoints that have been proposed for use in emerging disease response may not be silver bullets: while some of the dynamic effects and uncertainty (e.g. whether trial participants will be infected or not) are removed from the problem, some uncertainty (e.g. about how non-clinical outcomes relate to clinical outcomes) is also added. These considerations open exciting avenues for further research.

Declarations of interest

None.

References

[1] Sophie Cousins. WHO hedges its bets: the next global pandemic could be disease X. BMJ, page k2015, may 2018. doi: 10.1136/bmj.k2015.

[2] World Health Organizatrion (WHO). Pathogens prioritization: a scientific framework for epidemic and pandemic research preparedness, July 2024. https://www.who.int/publications/m/item/pathogens-prioritization-a-scientific-framework-for-epidemic-and-pandemic-research-preparedness.

[3] Gerardo Chowell and Cécile Viboud. Ebola vaccine trials: a race against the clock. The Lancet Infectious Diseases, 15(6):624–626, jun 2015. doi: 10.1016/s1473-3099(15)70159-3.

[4] Lori E Dodd. Clinical trials during the ebola crisis. Clinical Trials, 13(1):5–5, jan 2016. doi: 10.1177/1740774515621509.

[5] Gary Humphreys and Lynn Eaton. Accelerating vaccine trials. Bulletin of the World Health Organization, 99(7):482–483, July 2021. ISSN 0042-9686. doi: 10.2471/blt.21.020721.

[6] Melanie Saville, Jakob P. Cramer, Matthew Downham, Adam Hacker, Nicole Lurie, Lieven Van der Veken, Mike Whelan, and Richard Hatchett. Delivering pandemic vaccines in 100 days — what will it take? New England Journal of Medicine, 387(2):e3, jul 2022. doi: 10.1056/nejmp2202669.

[7] Marie Mura, Aurélie Trignol, Erwan Le Dault, and Jean-Nicolas Tournier. Lessons for medical countermeasure development from unforeseen outbreaks. Emerging Microbes & Infections, 14(1), March 2025. ISSN 2222-1751. doi: 10.1080/22221751.2025.2471035.

[8] Fergus Sweeney. EU's regulatory process for evaluationand approval of vaccines. European Medicine Agency (EMA), December 2020.

[9] U.S. Food and Drug Administration (FDA). Vaccine Development – 101, December 2020. https://www.fda.gov/vaccines-blood-biologics/development-approval-process-cber/vaccine-development-101.

[10] World Health Organization et al. Design of vaccine efficacy trials to be used during public health emergencies—points of considerations and key principles, 2020.

[11] Jozef Nauta. Statistics in Clinical and Observational Vaccine Studies. Springer International Publishing, 2020. ISBN 9783030376932. doi: 10.1007/978-3-030-37693-2.

[12] Steven Black, Juhani Eskola, Claire-Anne Siegrist, Neal Halsey, Noni MacDonald, Barbara Law, Elizabeth Miller, Nick Andrews, Julia Stowe, Daniel Salmon, Kirsten Vannice, Hector S Izurieta, Aysha Akhtar, Mike Gold, Gabriel Oselka, Patrick Zuber, Dina Pfeifer, and Claudia Vellozzi. Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines. The Lancet, 374(9707):2115–2122, dec 2009. doi: 10.1016/s0140-6736(09)61877-8.

[13] Steven B. Black, Barbara Law, Robert T. Chen, Cornelia L. Dekker, Miriam Sturkenboom, Wan-Ting Huang, Marc Gurwith, and Greg Poland. The critical role of background rates of possible adverse events in the assessment of covid-19 vaccine safety. Vaccine, 39(19):2712–2718, May 2021. ISSN 0264-410X. doi: 10.1016/j.vaccine.2021.03.016.

[14] Jim Weatherall, Faisal M. Khan, Mishal Patel, Richard Dearden, Khader Shameer, Glynn Dennis, Gabriela Feldberg, Thomas White, and Sajan Khosla. Clinical trials, real-world evidence, and digital medicine, pages 191–215. Elsevier, 2021. ISBN 9780128200452. doi: 10.1016/b978-0-12-820045-2.00011-8.

[15] U.S. Food and Drug Administration (FDA). Biological Approvals by Year. `https://www.fda.gov/vaccines-blood-biologics/development-approval-process-cber/biological-approvals-year`, 2025.

[16] Deborah F. King, Helen Groves, and Charlotte Weller. The role of correlates of protection in overcoming barriers to vaccine development and demonstrating efficacy. npj Vaccines, 9(1), April 2024. ISSN 2059-0105. doi: 10.1038/s41541-024-00873-5.

[17] A. Wilder-Smith, I. Longini, P. L. Zuber, T. Bärnighausen, W. J. Edmunds, N. Dean, V. Masserey Spicher, M. R. Benissa, and B. D. Gessner. The public health value of vaccines beyond efficacy: methods, measures and outcomes. BMC Medicine, 15(1), July 2017. ISSN 1741-7015. doi: 10.1186/s12916-017-0911-8.

[18] Kevin Bowrin, Jean-Baptiste Briere, Pierre Levy, Aurélie Millier, Emilie Clay, and Mondher Toumi. Cost-effectiveness analyses using real-world data: an overview of the literature. Journal of Medical Economics, 22(6):545–553, March 2019. ISSN 1941-837X. doi: 10.1080/13696998.2019.1588737.

[19] John Clemens. Evaluating New Vaccines for Developing Countries: Efficacy or Effectiveness? JAMA, 275(5):390, February 1996. ISSN 0098-7484. doi: 10.1001/jama.1996.03530290060038.

[20] Brian E. Rittenhouse. Exorcising protocol-induced spirits:: Making the clinical trial relevant for economics. Medical Decision Making, 17(3):331–339, July 1997. ISSN 1552-681X. doi: 10.1177/0272989x9701700310.

[21] John B. Kostis and Jeanne M. Dobrzynski. Limitations of randomized clinical trials. The American Journal of Cardiology, 129:109–115, August 2020. ISSN 0002-9149. doi: 10.1016/j.amjcard.2020.05.011.

[22] Mitra Saadatian-Elahi, Olaf Horstick, Robert F. Breiman, Bradford D. Gessner, Duane J. Gubler, Jacques Louis, Umesh D. Parashar, Roberto Tapia, Valentina Picot, Jean-Antoine Zinsou, and Christopher B. Nelson. Beyond efficacy: The full public health impact of vaccines. Vaccine, 34(9):1139–1147, February 2016. ISSN 0264-410X. doi: 10.1016/j.vaccine.2016.01.021.

[23] Bradford D. Gessner, David Kaslow, Jacques Louis, Kathleen Neuzil, Katherine L. O'Brien, Valentina Picot, Tikki Pang, Umesh D. Parashar, Mitra Saadatian-Elahi, and Christopher B. Nelson. Estimating the full public health value of vaccination. Vaccine, 35(46):6255–6263, November 2017. ISSN 0264-410X. doi: 10.1016/j.vaccine.2017.09.048.

[24] Amy K. O'Sullivan, David Thompson, and Michael F. Drummond. Collection of Health-Economic Data Alongside Clinical Trials: Is There a Future for Piggyback Evaluations? Value in Health, 8(1):67–79, January 2005. ISSN 1098-3015. doi: 10.1111/j.1524-4733.2005.03065.x.

[25] S. Petrou and A. Gray. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. BMJ, 342(apr07 2):d1548–d1548, April 2011. ISSN 1468-5833. doi: 10.1136/bmj.d1548.

[26] Scott D. Ramsey, Richard J. Willke, Henry Glick, Shelby D. Reed, Federico Augustovski, Bengt Jonsson, Andrew Briggs, and Sean D. Sullivan. Cost-effectiveness analysis alongside clinical trials ii—an ispor good research practices task force report. Value in Health, 18(2):161–172, March 2015. ISSN 1098-3015. doi: 10.1016/j.jval.2015.02.001.

[27] Mohamed El Alili, Johanna M. van Dongen, Jonas L. Esser, Martijn W. Heymans, Maurits W. van Tulder, and Judith E. Bosmans. A scoping review of statistical methods for trial-based economic evaluations: The current state of play. Health Economics, 31(12):2680–2699, September 2022. ISSN 1099-1050. doi: 10.1002/hec.4603.

[28] Angela Jornada Ben, Johanna M. van Dongen, Mohamed El Alili, Jonas L. Esser, Hana Marie Broulikova, and Judith E. Bosmans. Conducting Trial-Based Economic Evaluations Using R: A Tutorial. PharmacoEconomics, 41(11):1403–1413, July 2023. ISSN 1179-2027. doi: 10.1007/s40273-023-01301-7.

[29] Jordan M. Mitchell and Julie A. Patterson. The Inclusion of Economic Endpoints as Outcomes in Clinical Trials Reported to ClinicalTrials.gov. Journal of Managed Care and Specialty Pharmacy, 26(4):386–393, April 2020. ISSN 2376-1032. doi: 10.18553/jmcp.2020.26.4.386.

[30] Kristin N Nelson, Gavin Churchyard, Frank Cobelens, Willem A Hanekom, Philip C Hill, Benjamin Lopman, Vidya Mave, Molebogeng X Rangaka, Johan Vekemans, Richard G White, Emily B Wong, Leonardo Martinez, and Alberto L García-Basteiro. Measuring indirect transmission-reducing effects in tuberculosis vaccine efficacy trials: why and how? The Lancet Microbe, 4(8):e651–e656, August 2023. ISSN 2666-5247. doi: 10.1016/s2666-5247(23)00112-x.

[31] Jon Zelner, Julien Riou, Ruth Etzioni, and Andrew Gelman. Accounting for uncertainty during a pandemic. Patterns, 2(8):100310, August 2021. ISSN 2666-3899. doi: 10.1016/j.patter.2021.100310.

[32] Julien Flaig and Nicolas Houy. Optimal epidemic control under uncertainty: Tradeoffs between information collection and other actions. Med. Decis. Making, 43(3):350–361, April 2023.

[33] Nicolas Houy and Julien Flaig. Value of information dynamics in Disease X vaccine clinical trials. Vaccine, 42(7):1521–1533, March 2024. ISSN 0264-410X. doi: 10.1016/j.vaccine.2024.01.063.

[34] Benjamin R Saville, Jason T Connor, Gregory D Ayers, and JoAnn Alvarez. The utility of bayesian predictive probabilities for interim monitoring of clinical trials. Clinical Trials, 11(4):485–493, May 2014. ISSN 1740-7753. doi: 10.1177/1740774514531352.

[35] Steven E Bellan, Juliet R C Pulliam, Carl A B Pearson, David Champredon, Spencer J Fox, Laura Skrip, Alison P Galvani, Manoj Gambhir, Ben A Lopman, Travis C Porco, Lauren Ancel Meyers, and Jonathan Dushoff. Statistical power and validity of ebola vaccine trials in sierra leone: a simulation study of trial design and analysis. The Lancet Infectious Diseases, 15(6):703–710, jun 2015. doi: 10.1016/s1473-3099(15)70139-8.

[36] Anton Camacho, Rosalind M Eggo, Sebastian Funk, Conall H Watson, Adam J Kucharski, and W John Edmunds. Estimating the probability of demonstrating vaccine efficacy in the declining ebola epidemic: a bayesian modelling approach. BMJ Open, 5(12):e009346, dec 2015. doi: 10.1136/bmjopen-2015-009346.

[37] An Vandebosch, Robin Mogg, Nele Goeyvaerts, Carla Truyers, Brian Greenwood, Debby Watson-Jones, Guillermo Herrera-Taracena, Wim Parys, and Tony Vangeneugden. Simulation-guided phase 3 trial design to evaluate vaccine effectiveness to prevent ebola virus disease infection: Statistical considerations, design rationale, and challenges. Clinical Trials, 13(1):57–65, jan 2016. doi: 10.1177/1740774515621059.

[38] A. Camacho, R.M. Eggo, N. Goeyvaerts, A. Vandebosch, R. Mogg, S. Funk, A.J. Kucharski, C.H. Watson, T. Vangeneugden, and W.J. Edmunds. Real-time dynamic modelling for the design of a cluster-randomized phase 3 ebola vaccine trial in sierra leone. Vaccine, 35(4):544–551, jan 2017. doi: 10.1016/j.vaccine.2016.12.019.

[39] Matt David Thomas Hitchings, Rebecca Freeman Grais, and Marc Lipsitch. Using simulation to aid trial design: Ring-vaccination trials. PLOS Neglected Tropical Diseases, 11(3):e0005470, mar 2017. doi: 10.1371/journal.pntd.0005470.

[40] Donald A. Berry, Scott Berry, Peter Hale, Leah Isakov, Andrew W. Lo, Kien Wei Siah, and Chi Heem Wong. A cost/benefit analysis of clinical trial designs for COVID-19 vaccine candidates. PLOS ONE, 15(12):e0244418, dec 2020. doi: 10.1371/journal.pone.0244418.

[41] Rob Johnson, Chris Jackson, Anne Presanis, Sofia S. Villar, and Daniela De Angelis. Quantifying efficiency gains of innovative designs of two-arm vaccine trials for COVID-19 using an epidemic simulation model. Statistics in Biopharmaceutical Research, 14(1):33–41, jul 2021. doi: 10.1080/19466315.2021.1939774.

[42] Johan L. Plas, Michiel J. Esdonk, Ingrid M. C. Kamerling, and Adam F. Cohen. Accelerating vaccine trial conduct in a pandemic with a hot spot-based inclusion

strategy using trial and epidemic simulation. Clinical and Translational Science, 14 (6):2391–2398, jul 2021. doi: 10.1111/cts.13104.

[43] Allan S. Detsky. Using cost-effectiveness analysis to improve the efficiency of allocating funds to clinical trials. Statistics in Medicine, 9(1-2):173–184, jan 1990. doi: 10.1002/sim.4780090124.

[44] Karl Claxton and John Posnett. An economic approach to clinical trial design and research priority-setting. Health Economics, 5(6):513–524, November 1996. ISSN 1099-1050. doi: 10.1002/(sici)1099-1050(199611)5:6<513::aid-hec237>3.0.co;2-9.

[45] Jose M. Bernardo. Statistical inference as a decision problem: the choice of sample size. Journal of the Royal Statistical Society: Series D (The Statistician), 46(2): 151–153, jul 1997. doi: 10.1111/1467-9884.00071.

[46] Dennis V. Lindley. The choice of sample size. Journal of the Royal Statistical Society: Series D (The Statistician), 46(2):129–138, jul 1997. doi: 10.1111/1467-9884.00068.

[47] S. B. Tan and A. F. M. Smith. Exploratory thoughts on clinical trials with utilities. Statistics in Medicine, 17(23):2771–2791, dec 1998. doi: 10.1002/(sici)1097-0258(19981215)17:23<2771::aid-sim42>3.0.co;2-9.

[48] A. O'Hagan and J. W. Stevens. Bayesian assessment of sample size for clinical trials of cost-effectiveness. Medical Decision Making, 21(3):219–230, May 2001. ISSN 0272-989X. doi: 10.1177/0272989x0102100307.

[49] Andrew R. Willan and Eleanor M. Pinto. The value of information and optimal clinical trial design. Statistics in Medicine, 24(12):1791–1806, 2005. ISSN 1097-0258. doi: 10.1002/sim.2069.

[50] Peter Bacchetti, Charles E McCulloch, and Mark R Segal. Simple, defensible sample sizes based on cost efficiency. Biometrics, 64(2):577–585, 2008. doi: 10.1111/j.1541-0420.2008.01004_1.x.

[51] Andrew R. Willan and Simon Eckermann. Optimal clinical trial design using value of information methods with imperfect implementation. Health Economics, 19(5): 549–561, April 2009. ISSN 1099-1050. doi: 10.1002/hec.1493.

[52] Bas Groot Koerkamp, Sandra Spronk, Theo Stijnen, and M.G. Myriam Hunink. Value of information analyses of economic randomized controlled trials: The treatment of

intermittent claudication. Value in Health, 13(2):242–250, March 2010. ISSN 1098-3015. doi: 10.1111/j.1524-4733.2009.00656.x.

[53] Natalia Muehlemann, Tianjian Zhou, Rajat Mukherjee, Munshi Imran Hossain, Satrajit Roychoudhury, and Estelle Russek-Cohen. A Tutorial on Modern Bayesian Methods in Clinical Trials. Therapeutic Innovation & Regulatory Science, 57(3):402–416, April 2023. ISSN 2168-4804. doi: 10.1007/s43441-023-00515-3.

[54] Stephen J. Ruberg, Francois Beckers, Rob Hemmings, Peter Honig, Telba Irony, Lisa LaVange, Grazyna Lieberman, James Mayne, and Richard Moscicki. Application of bayesian approaches in drug development: starting a virtuous cycle. Nature Reviews Drug Discovery, 22(3):235–250, February 2023. ISSN 1474-1784. doi: 10.1038/s41573-023-00638-0.

[55] Mohsen Sadatsafavi, Carlo Marra, Shawn Aaron, and Stirling Bryan. Incorporating external evidence in trial-based cost-effectiveness analyses: the use of resampling methods. Trials, 15(1), June 2014. ISSN 1745-6215. doi: 10.1186/1745-6215-15-201.

[56] Nicky J Welton, Mark Strong, Christopher Jackson, and Gianluca Baio. Economic evaluation and cost-effectiveness of health care interventions. In Bayesian Methods in Pharmaceutical Research, pages 301–315. Chapman and Hall/CRC, 2020.

[57] Dalene Stangl. The Use of Reference Priors and Bayes Factors in the Analysis of Clinical Trials, pages 237–249. Springer New York, 2000. ISBN 9781461212843. doi: 10.1007/978-1-4612-1284-3_7.

[58] James M. Brophy. Placing trials in context using bayesian analysis: Gusto revisited by reverend bayes. JAMA, 273(11):871, March 1995. ISSN 0098-7484. doi: 10.1001/jama.1995.03520350053028.

[59] European Medicine Agency (EMA). EMA considerations on COVID-19 vaccine approval – Scientific guideline, November 2020. `https://www.ema.europa.eu/en/ema-considerations-covid-19-vaccine-approval-scientific-guideline`.

[60] U.S. Food and Drug Administration (FDA). Emergency Use Authorization for Vaccines to Prevent COVID-19 – Guidance for Industry, February 2021.

[61] World Health Organizatrion (WHO). Considerations for the Assessment of COVID-19 Vaccines for Listing by WHO, November 2020. `https://www.who.int/publications/m/item/considerations-for-the-assessment-of-covid-19-vaccines-for-listing-by-who`.

[62] Steven A. Frost, Evan Alexandrou, Luis Schulz, and Anders Aneman. Interpreting the results of clinical trials, embracing uncertainty: A Bayesian approach. Acta Anaesthesiologica Scandinavica, 65(2):146–150, October 2020. ISSN 1399-6576. doi: 10.1111/aas.13725.

[63] Daniel Hill-McManus and Dyfrig A. Hughes. Combining Model-Based Clinical Trial Simulation, Pharmacoeconomics, and Value of Information to Optimize Trial Design. CPT: Pharmacometrics and Systems Pharmacology, 10(1):75–83, December 2020. ISSN 2163-8306. doi: 10.1002/psp4.12579.

[64] Geoffrey A. Weinberg and Peter G. Szilagyi. Vaccine epidemiology: Efficacy, effectiveness, and the translational research roadmap. The Journal of Infectious Diseases, 201(11):1607–1610, June 2010. ISSN 1537-6613. doi: 10.1086/652404.

[65] Fernando P. Polack, Stephen J. Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L. Perez, Gonzalo Pérez Marc, Edson D. Moreira, Cristiano Zerbini, Ruth Bailey, Kena A. Swanson, Satrajit Roychoudhury, Kenneth Koury, Ping Li, Warren V. Kalina, David Cooper, Robert W. Frenck, Laura L. Hammitt, Özlem Türeci, Haylene Nell, Axel Schaefer, Serhat Ünal, Dina B. Tresnan, Susan Mather, Philip R. Dormitzer, Uğur Şahin, Kathrin U. Jansen, and William C. Gruber. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. New England Journal of Medicine, 383(27):2603–2615, December 2020. ISSN 1533-4406. doi: 10.1056/nejmoa2034577.

[66] Anthony O'Hagan, John W. Stevens, and Michael J. Campbell. Assurance in clinical trial design. Pharmaceutical Statistics, 4(3):187–201, 2005. ISSN 1539-1612. doi: 10.1002/pst.175.

[67] David Ferreira, Pierre-Olivier Ludes, Pierre Diemunsch, Eric Noll, Klaus D. Torp, and Nicolas Meyer. Bayesian predictive probabilities: a good way to monitor clinical trials. British Journal of Anaesthesia, 126(2):550–555, February 2021. ISSN 0007-0912. doi: 10.1016/j.bja.2020.08.062.

[68] Holger Kraiczi and Marianne Frisén. Effect of uncertainty about population parameters on pharmacodynamics-based prediction of clinical trial power. Contemporary Clinical Trials, 26(2):118–130, April 2005. ISSN 1551-7144. doi: 10.1016/j.cct.2004.10.002.

[69] Yuanyuan Cui, Brendan Murphy, Anastasia Gentilcore, Yugal Sharma, Lori M Minasian, Barnett S Kramer, Paul M Coates, John K Gohagan, Juergen Klenk, and

Bruce Tidor. Multilevel modeling and value of information in clinical trial decision support. BMC Systems Biology, 8(1), December 2014. ISSN 1752-0509. doi: 10.1186/s12918-014-0140-0.

[70] T. Alex Perkins, Robert C. Reiner, Guido España, Quirine A. ten Bosch, Amit Verma, Kelly A. Liebman, Valerie A. Paz-Soldan, John P. Elder, Amy C. Morrison, Steven T. Stoddard, Uriel Kitron, Gonzalo M. Vazquez-Prokopec, Thomas W. Scott, and David L. Smith. An agent-based model of dengue virus transmission shows how multiple uncertainties about vaccine efficacy influence public health impact projections. bioRxiv, October 2016. doi: 10.1101/082396.

[71] Winston E Abara, Julianne Gee, Mark Delorey, Ye Tun, Yi Mu, David K Shay, and Tom Shimabukuro. Expected rates of select adverse events after immunization for coronavirus disease 2019 vaccine safety monitoring. The Journal of Infectious Diseases, 225(9):1569–1574, dec 2021. doi: 10.1093/infdis/jiab628.

[72] Diane Gubernot, Amelia Jazwa, Manette Niu, Jane Baumblatt, Julianne Gee, Pedro Moro, Jonathan Duffy, Theresa Harrington, Michael M. McNeil, Karen Broder, John Su, Satoshi Kamidani, Christine K. Olson, Lakshmi Panagiotakopoulos, Tom Shimabukuro, Richard Forshee, Steven Anderson, and Sarah Bennett. U.s. population-based background incidence rates of medical conditions for use in safety assessment of COVID-19 vaccines. Vaccine, 39(28):3666–3677, jun 2021. doi: 10.1016/j.vaccine.2021.05.016.

[73] HRSA. Vaccine Injury Compensation Data. `https://www.hrsa.gov/vaccine-compensation/data`, 2023. Accessed: 2023-05-10.

[74] Vaccine Damage Payment. `https://www.gov.uk/vaccine-damage-payment`, 2023. Accessed: 2023-05-10.

[75] OECD. Average wages. `https://data.oecd.org/earnwage/average-wages.htm`, 2023. Accessed: 2023-05-10.

[76] CDC Vaccine Price List. `https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/price-list/index.html`, 2023. Accessed: 2023-05-10.

[77] S. Mokiou, B. Standaert, X. Li, and E. De Cock. Measuring the cost of a pediatric vaccine administration in the UK. Vaccine, 36(2):237–242, jan 2018. doi: 10.1016/j.vaccine.2017.11.042.

[78] ICH. E9 Statistical principles for clinical trials. Technical report, European Medicines Agency (EMA), September 1998. URL https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline#current-version-section. CPMP/ICH/363/96.

[79] Anne McNicholas, Yvonne Galloway, Paul Stehr-Green, Stewart Reid, Sarah Radke, Kerry Sexton, Charlotte Kieft, Claire Macdonald, Jocelyn Neutze, Ross Drake, Dorothy Isaac, Mary O'Donnell, Michael Tatley, Philipp Oster, and Jane O'Hallahan. Post-Marketing Safety Monitoring of a New Group B Meningococcal Vaccine in New Zealand, 2004-2006. Human Vaccines, 3(5):196–204, sep 2007. doi: 10.4161/hv.3.5.4458.

[80] Helen Petousis-Harris. Assessing the safety of COVID-19 vaccines: A primer. Drug Safety, 43(12):1205–1210, sep 2020. doi: 10.1007/s40264-020-01002-6.

A Additional figure: transmission model

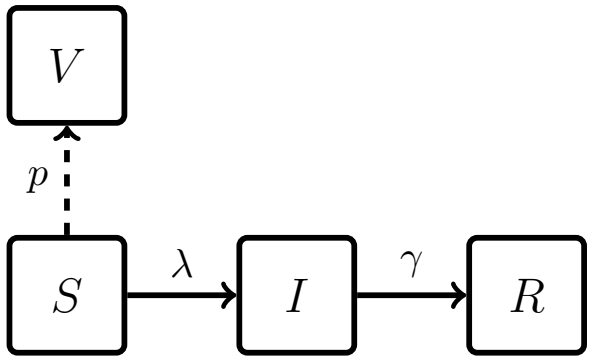
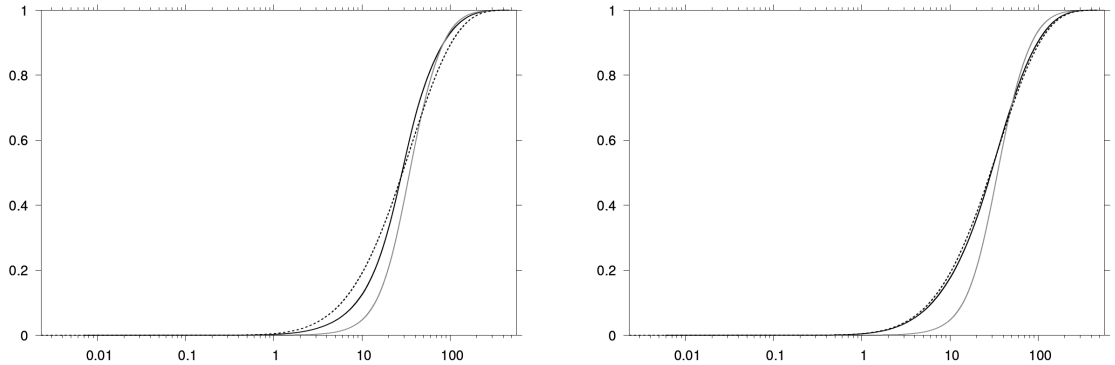


Figure App-1: Sketch of the transmission model. Dashed: possible vaccination (instantaneous). $\lambda = \gamma R_0 \frac{I}{N}$ is the force of infection.

B Additional analyses: baseline scenario

Figure App-2 shows total cost cumulative distributions obtained in our baseline scenario (N_{ct} = 50,000, T_{ct} = 80 days). The curve for "trial, vaccination at $T_v + T_{ct}$" cannot be distinguished from the curve for "no trial, vaccination at T_v". The curve for "no trial, no vaccination" cannot be distinguished from the curve for "trial, no vaccination".



(a) CT0 (b) CT95

Figure App-2: Total cost (10^9 USD) cumulative distributions for vaccination policies: clinical trials CT0 and CT95 (black), vaccination without trial "no trial, vaccination at T_v" (grey) and no vaccination with trial "trial, no vaccination" (dashed black). Curves obtained from 400,000 parameter draws.

Table App-1 shows estimated cost assuming that all uncertain parameters (including

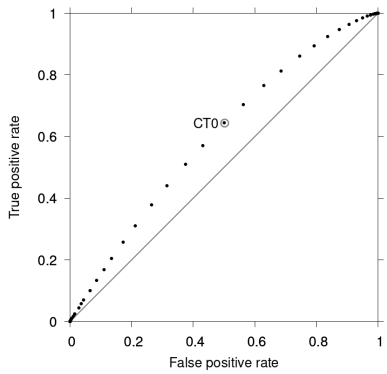
efficacy and adverse event parameters) are equal to their prior expected value. Under this assumption, the best policy is to do nothing. Implementing a clinical trial increases the total cost through vaccination costs and adverse events among the $N_{ct}/2$ vaccinated individuals. After a trial, at time $T_v + T_{ct}$, the best policy is to not vaccinate. Notice that we do not use clinical trial information here, and that policies "trial, optimal at $T_v + T_{ct}$", "CT0" and "CT95" are trivially equivalent.

	Cost			
Policy	**Infections**	**Vaccination**	**Adverse events**	**Total**
No trial, no vaccination	28.34	0	0	28.34
No trial, vaccination at T_v	26.34	9.34	13.84	49.52
Trial, no vaccination	28.34	3.22×10^{-3}	4.92×10^{-3}	28.35
Trial, vaccination at $T_v + T_{ct}$	27.84	9.11	13.49	50.43
Trial, optimal at $T_v + T_{ct}$	28.34	3.22×10^{-3}	4.92×10^{-3}	28.35
CT0	28.34	3.22×10^{-3}	4.92×10^{-3}	28.35
CT95	28.34	3.22×10^{-3}	4.92×10^{-3}	28.35

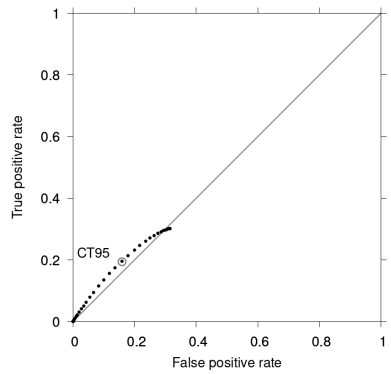
Table App-1: Summary simulation results for different vaccination policies, with uncertain parameters set to their prior expected value. Costs in 10^9 USD.

C RCT performance in the ROC plane

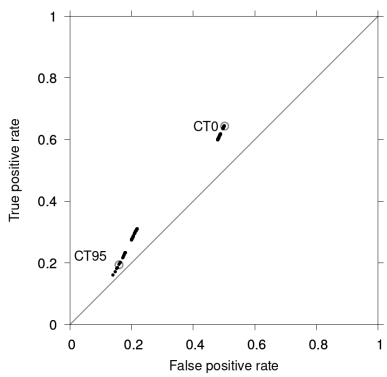
The performance of clinical trials can also be illustrated with receiver operating characteristic (ROC) curves. In Figure App-3, we display the performance obtained in the ROC plane by varying safety threshold λ_{ae}^{max}, efficacy confidence level $1 - \alpha$, or both. The simulations are the same as in Figures 2, 3, and 4 respectively. The true positive rate is the proportion of simulations for which the vaccine passes both tests, among the simulations for which it is optimal to vaccinate at time $T_v + T_{ct}$ following the trial. The false positive rate is the proportion of simulations for which the vaccine passes both tests, among the simulations for which it is optimal to not vaccinate at time $T_v + T_{ct}$. Recall that the true positive and false positive rates of CT0 and CT95 can be retrieved from Table 3 as shown in Section 3.1. RCTs in the $(0, 0)$ corner are too restrictive, RCTs in the $(1, 1)$ corner are too permissive, and the $(0, 1)$ corner corresponds to a perfect classifier. For some permissive safety thresholds, the RCT performs worse than random guess (dots below the line $y = x$ in Figure (App-3d)).



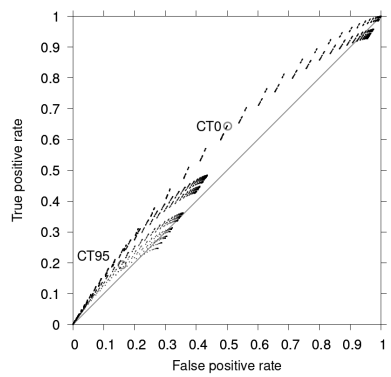
(a) CT0, varying λ_{ae}^{max}



(b) CT95, varying λ_{ae}^{max}



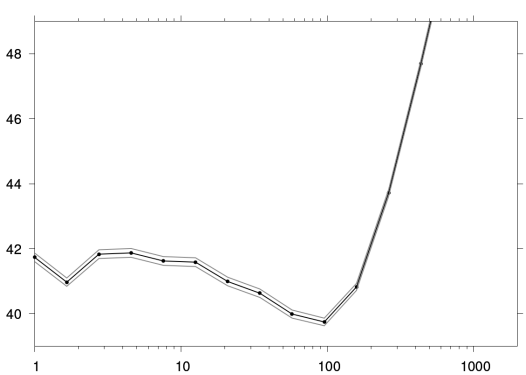
(c) Varying $1 - \alpha$



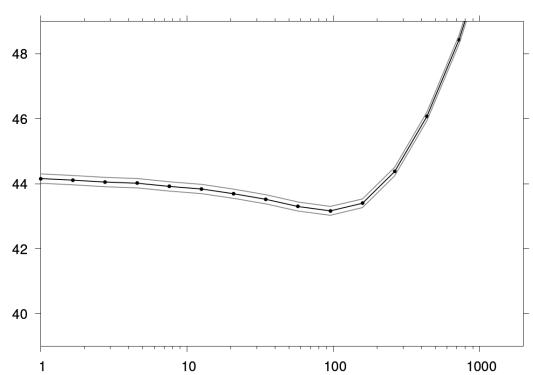
(d) Varying $1 - \alpha$ and λ_{ae}^{max}

Figure App-3: ROC curves obtained by varying safety threshold λ_{ae}^{max}, efficacy confidence level $1 - \alpha$, or both. Grey circles: CT0 and CT95.

D Additional figure: varying T_{ct}



(a) CT0



(b) CT95

Figure App-4: Total cost (10^9 USD) as a function of the duration T_{ct} of the clinical trial (days) for policies "CT0" and "CT95". Grey: 95% CI.