This is a repository copy of *The design and analysis of vaccine trials for COVID-19 for the purpose of estimating efficacy*.

SPECIAL ISSUE PAPER

WILEY

The design and analysis of vaccine trials for COVID-19 for the purpose of estimating efficacy

Stephen Senn

Medical Statistics Group, School of Health and Related Research, University of Sheffield, Sheffield, UK

Correspondence
Stephen Senn, Medical Statistics Group, School of Health and Related Research, University of Sheffield, Sheffield, UK.
Email: stephen@senns.uk

Abstract

After a preliminary explanation as to how I came to know Andy Grieve and some remarks about his career and mine and how they have intersected, I consider the design and analysis of trials of vaccines for COVID-19 for the purpose of estimating efficacy. Five large trials, run by the sponsors Pfizer/BioNTech, AstraZeneca/Oxford University, Moderna, Novavax and J&J Janssen are considered briefly. Frequentist approaches to analysis were used for four of the trials but Pfizer/BioNTech nominated a Bayesian approach. The design and analysis of this trial is considered in some detail, in particular as regards the choice of prior distribution. I conclude by drawing some general lessons.

KEYWORDS
Bayesian, conditional inference, design of experiments, frequentist, sequential design, vaccine efficacy

1 | INTRODUCTION

It is a pleasure to be able to contribute a piece to Andy Grieve's festschrift issue. A short explanation as to how I have to know Andy is in order. In 1987, after 9 years working for the then Dundee College of Technology, I was appointed as a statistician in my hometown of Basle to work in the medical statistics group of CIBA-Geigy (now merged with Sandoz to form Novartis) headed by Jakob Schenker. There was another, methodological statistics group headed by Hugo Flühler. The medical statistics group was located in the Klybeck site and the methodology group in the Rosental site, over a mile away and this meant that the statisticians from the two groups did not meet up regularly, although they did occasionally attend seminars together. However, my boss Jakob Schenker previously worked in the methodology group and brought some statisticians with him when he took over the medical statistics group and their familiarity with the other group helped the medical statisticians to know what their more methodologically minded colleagues were up to.

Nevertheless, I found this separation far from ideal and remarked that CIBA-Geigy seemed to have two statistics groups: one with lots to do and no time to think and the other with lots of time to think and nothing to do. This was, however, unfair. The methodology group was, in fact, involved in many applications, basically anything to do with statistics except phase II–IV clinical trials, and they were also involved across all the many divisions of the conglomerate,

not just pharmaceuticals. They had been pioneering practical Bayesian approaches to a number of problems they encountered. An important context for this is that this was before the computational revolution introduced at the beginning of the next decade by Alan Gelfand and Adrian Smith.[1]

In fact, only a year earlier, in collaboration with Adrian Smith, who acted as a statistical consultant to the group, Hugo Flühler and two members of his group had presented a read paper to the Royal Statistical Society.[2] The two others were Amy Racine and Andy Grieve and their names regularly cropped up in our discussions of what the methodology group was up to.

I was advised on my arrival to become familiar with SAS® and also repeated measures designs and was informed that the local expert for the latter was Andy Grieve. He had a fearsome reputation as someone who could master detail easily and 5 years earlier had published a note in Biometrics[3] correcting some algebra in an earlier famous paper by James Grizzle on crossover designs.[4] According to local legend, whenever Andy found a problem difficult he would just give himself leave of absence from work and then emerge a few days later with the problem solved. I was given a technical report of his to read that I found intimidatingly stiff with algebra (some of it related no-doubt, to his work on sphericity[5] and pre-test post-test designs[6]) and immediately became completely in awe of his statistical ability, a feeling that has not left me since.

In fact, unfortunately for me, Andy was about to move to the English office of CIBA-Geigy, so that we hardly overlapped in Basle. However, his new appointment in the UK meant that he was more often involved in clinical trials and we collaborated on a project in asthma developing the drug formoterol,[7] so we came to know each other well by phone (this was an era before email for us). I soon came to count Andy as a good friend as well as a reliable source of sound statistical advice and we have collaborated several times since, having published together on at least five occasions. A highlight of our collaboration for me was having appeared together as two of the four members of the Royal Statistical Society team that got to the final of the 2006 *University Challenge the Professionals* series.

Andy is a committed Bayesian and I often fly under a frequentist flag of convenience. This has led to a number of good-natured arguments over the years between the two of us, although we have also often agreed that much of what matters has little to do with statistical philosophy per se but rather more with understanding and respecting the field of application. An example is given by our joint paper on crossover trials.[8] I have decided to pick as my topic for this festschrift issue something that ought to interest us all, the design of trials of efficacy of vaccines against COVID, but which also gives scope for a double connection to Andy. I shall consider five trials but only one of them in detail. This was not only a trial run by Pfizer, a company for whom Andy used to work but also one for which Bayesian methods were used. Andy had long left Pfizer by the time the vaccine development programme for COVID was put together and he is not responsible for the decisions made in planning this trial but I like to think that something of the Bayesian approach he brought survives, so this seems to me to be a suitable topic for discussion.

2 | FIVE TRIALS

The five trials that I shall consider are all large, randomised placebo-controlled phase III studies and are summarised in Table 1. I shall only consider the trials as regards the purpose of estimating efficacy. Safety is also very important but

TABLE 1 Numbers of subjects and cases for five large trials.

Sponsor	Subjects		Cases	
	Vaccine	Placebo	Vaccine	Placebo
Pfizer/BioNTech[9]	20,712	21,096	77	850
AZ/Oxford[10]	17,662	8550	73	130
Moderna[11]	14,134	14,073	11	185
Novavax[12]	7020	7019	10	96
J&J Janssen[13]	19,630	19,691	116	348

Note: Pfizer/BioNTech figures after 6 months follow up. Novavax after 3 months follow up AZ/Oxford, Moderna J&J Janssen after 2 months follow-up (figures are the median time after the second dose except for J&J Janssen, where time is after the first dose).

TABLE 2 Endpoints used for the primary efficacy analysis of the five trials.

Sponsor	Design	Endpoint
Pfizer/BioNTech	Observer blind	Covid-19 occurrence 7+ days after the 2nd dose in participants without evidence of infection
AZ/Oxford	Double blind	1st occurrence of SARS-CoV-2 RT-PCR confirmed symptomatic illness, 15+ days after 2nd dose
Moderna	Observer blind	1st occurrence of symptomatic Covid-19 with onset 14+ days after the 2nd injection
Novavax	Observer blind	1st occurrence of PCR-confirmed symptomatic (mild to severe) COVID-19 with onset 7+ days after 2nd vaccine
J&J Janssen	Double blind	Moderate to severe critical COVID-19 occurring 14+ days after vaccination

TABLE 3 Various features of the plans for the five trials.

| Sponsor | Vaccine Efficacy | | Assumed Event rate | | Target numbers | | |
	H0%	H1%	% per 6 months	Power (%)	Events	Vaccine Subjects	Placebo Subjects
Pfizer/BioNTech[9]	30	60	0.65	90	164	21,999	21,999
AZ/Oxford[10]	30	60	0.8	90	150	20,000	10,000
Moderna[11]	30	60	0.7	90	151	15,000	15,000
Novavax[12]	30	70	1	95	100	7500	7500
J&J Janssen[13]	30	60	0.7	90	154	30,000	30,000

raises different issues as regards the use of controls and also concerning scales of measurement.[14] Such matters are beyond the scope of this article but are either applied or discussed in various papers to which I refer.

Attempting to describe the COVID vaccine trials during the pandemic has been trying to hit a moving target. The trials have reported at stages at which information is still accruing, even if earlier analyses presented were at the pre-specified primary time point, whether for accrual or follow-up. The reader should bear this in mind. What is being presented is a snapshot. I cannot guarantee, even at the time of writing, that what is presented is the latest picture of a given trial. It is the nature of the studies that the situation may change over time as more observation time accrues. The footnotes to the table should be consulted for details of follow up.

There were also some differences in the approaches to blinding and in the endpoint used and a brief summary is provided in Table 2. It should be noted that the J&J Janssen vaccine was given in a single injection, whereas for the other four sponsors, two injections at various intervals were given. For these four sponsors, only cases arising after the second injection were counted. The subsequent interval required to be counted varied from 7 days for Pfizer/BioNTech and Novavax to 14 and 15 for Moderna and AZ/Oxford.

Although it has little direct relevance to inference, it is of interest to look at the plans for the five trials. Details are summarised in Table 3. All the five trials used as a null hypothesis that the vaccine efficacy (VE) was 30%, suggesting either strong agreement in the community of researchers or common regulatory advice or both. Four sponsors, Pfizer/BioNTech, AZ/Oxford, Moderna and J & J Janssen used a clinically relevant value of 60% for VE for the power calculation. Novavax used a value of 70%. (Details of the VE scale will be discussed in due course).

Four of the trials used a one-to-one randomisation, whereas the AZ/Oxford trial allocated twice as many subjects to vaccine as to placebo. This was a surprising choice. For a given number of patients enrolled it would give more information on side-effects, albeit only in an uncontrolled manner. However, all the trials were sequential, and since stopping would be largely dependent on the number of cases, in the event of the vaccine's being effective, this would delay the point at which efficacy could be declared. Sequential boundaries for four of the trials are given in Figure 1. The exception is the J&J Janssen trial, which used a 'truncated sequential probability ratio test'[13] (p. 2189). It is noticeable, for the four trials using a VE of 60% under H_1, despite many other details varying, similar numbers of events are targeted, ranging from 150 for AZ/Oxford to 164 for Pfizer BioNTech. Novavax assumed a VE of 70% and targeted a much lower number of events, 100. See Patterson et al (2022), in particular their Table 1, for a useful summary of sample size considerations for vaccine trials.[15] In fact, the major uncertainty in such trials is not the number of events that are needed but the number of subjects needed to provide them. This is one reason that makes a sequential approach particularly attractive.
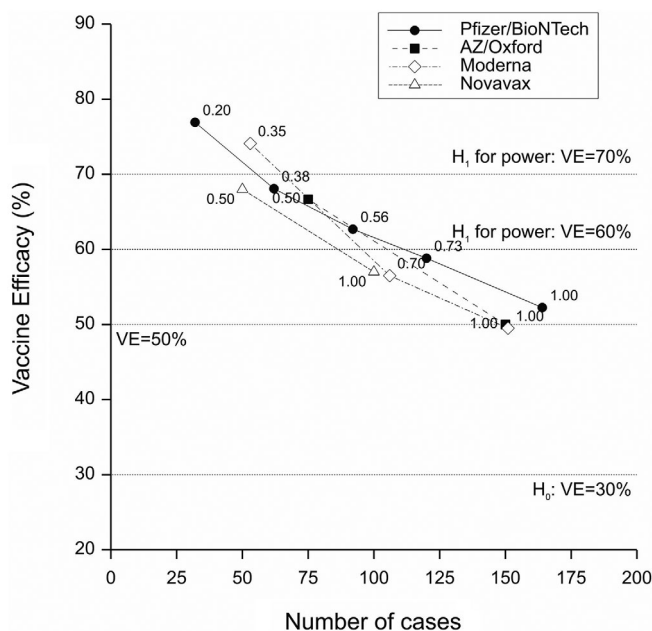
FIGURE 1 Stopping boundaries for four of the five vaccine efficacy trials. The numbers are the information fractions at the various looks.

Explicit details of the way the J&J Janssen trial was conducted as regards monitoring are neither given in the paper publishing the results nor in the protocol but reference is made to two papers[16,17] by statisticians working for GlaxoSmithKline (GSK), about which I can make three comments. First, it is pleasing to see ideas developed by statisticians at one pharmaceutical company (GSK) being used at another (J&J Janssen). (Vlad Dragalin moved from GSK to J&J Janssen so this is no doubt part of the explanation!) This is a tribute to the science of drug development and the pharmaceutical industry. Second, it is, however, a pity that specific details of implementation were omitted from the protocol and the publication. Third, the conditional analysis described by the GSK statisticians[16,17] is important and indeed central to much of what I shall subsequently describe. I have frequently used the conditional argument in blogging on COVID vaccine trials without realising that it had been very nicely described in these papers.[16,17] I am now in a position to make amends by citing them.

As regards the other boundaries, in some cases I have had to construct these from incomplete information in the trial protocols. The reader should beware it is possible that I have made mistakes. Novavax took one interim look using a Pocock boundary.[18] The Pfizer analysis was Bayesian and will be discussed in detail in due course but the boundary, according to the protocol, was chosen to control the type I error rate at 2.5% one sided and to allow for up to four interim looks. There was also a futility boundary, which will not be discussed here. AZ/Oxford planned one interim look at about 50% of information with an alpha level of 0.31% with 4.9% at the end for an overall type I error rate of 5%. Reference is made to a Lan-DeMets spending function[19] but the rule itself is not given a name in the protocol. The values are *fairly* similar to those for an O'Brien-Fleming rule.[20] Moderna planned three looks in total and make explicit reference to Lan-DeMets and O'Brien-Fleming.

It turns out, however, that the sequential approaches used have very little impact on the final analysis of the results of these trials. A possible approach to analysis will be discussed in the next section.

3 | VACCINE EFFICACY

The common scale used to report on the trials was vaccine efficacy, VE, defined as a parameter as

$$\text{VE} = 100\frac{\pi_\text{P} - \pi_\text{V}}{\pi_\text{P}} = 100\left(1 - \frac{\pi_\text{V}}{\pi_\text{P}}\right), \tag{1}$$

where π_P, π_V are the probabilities of infection if given placebo or vaccine and the multiplier 100 is used because the figure is usually expressed as a percentage. As the second of the forms shows, but for the factor 100, it is simply one minus the relative risk.[15] A simple estimate of this is

$$\mathrm{VE}_{\mathrm{est}} = 100 \frac{Y_P/n_P - Y_V/n_V}{Y_P/n_P}, \tag{2}$$

where Y_P, Y_V, n_P, n_V are numbers of cases of COVID and subjects treated in vaccine and placebo groups respectively. If $n_V = n_P$ then (2) reduces to

$$\mathrm{VE}_{\mathrm{est}} = 100 \frac{Y_P - Y_V}{Y_P}, \tag{3}$$

A simple justification of (3) is then as follows. (1) The numerator estimates the number of cases of COVID that were prevented: Y_V is the number of observed cases in the vaccine group but if the placebo group is similar in size and nature, then Y_P is an estimate of the counterfactual number of cases that would have been seen had the vaccine been ineffective. Thus $Y_P - Y_V$ is an estimate of the number of cases prevented. (2) The denominator that is used is now an estimate of the number of subjects at risk. The number of subjects given vaccine, n_V, might seem to be appropriate here but in fact, it is the nature of a pandemic that during any period of observation many individuals will not come into contact with the virus and therefore cannot be infected. They thus also cannot be protected from infection. Therefore, the number that is used is the number who would have been infected had the vaccine been ineffective and this is estimated as Y_P. More generally, if the number of subjects in the two groups are unequal, which was a design feature of the AZ/Oxford trial, then if $R = n_V/n_P$ one can write (2) as

$$\mathrm{VE}_{\mathrm{est}} = 100 \frac{RY_P - Y_V}{RY_P}. \tag{4}$$

Note that (3) is a special case of (4) with $R = 1$. Expression (4) is useful in that if numbers of subject are unknown but the allocation ratio is known, then, for a large trial, one can simply substitute the allocation ratio for R in (4) and calculate accordingly. This was the situation I found myself in in various stages in the pandemic when I was blogging on the results of various trials for which numbers of cases were usually available, say from press-releases, but numbers of subjects often were not. In fact, for modelling, a more useful parameter is what might be called the case proportion.

$$\theta = \frac{\pi_V}{\pi_P + \pi_V}. \tag{5}$$

Take for simplicity the balanced case with one to one randomization and therefore $R = 1$. If the number of cases in the placebo and vaccine groups are approximately Poisson distributed with parameters λ_P, λ_V, respectively, then

$$\begin{aligned} & Y_V \tilde{\mathrm{Bin}}(\theta, n) \\ & \theta = \frac{\lambda_V}{\lambda_P + \lambda_V}, n = Y_V + Y_P. \end{aligned} \tag{6}$$

which is to say that conditionally on the total number of cases, the number of cases under vaccine has a binomial distribution (θ, n) with θ equal to the expected number of cases under vaccine as a ratio of total expected cases and n equal to the number of cases.

    This form of conditioning is discussed in section 4.5 of Lehmann's famous book[21] and also by Dragalin, Fedorov[16] and in an earlier paper together with Cheuvart,[17] both of which papers describe its application to vaccine trials. An analogous form is, of course, common in log-linear models (see McCullagh and Nelder[22] chapter 6, in particular section 6.4). For a more recent discussion, see the paper by Patterson et al already referenced.[15] This particular form of analysis is very flexible, permits use of frequentist 'exact' methods for constructing confidence intervals and was employed by Pfizer/BioNTech in their Bayesian approach, as I shall discuss in due course.

The relationship between (1) and (5) is given by

$$VE = 100 \frac{2\theta - 1}{\theta - 1}$$
$$\theta = \frac{1 - VE/100}{2 - VE/100}.$$

(7)

This can be used to transfer inference from one scale to another.

4 | RESULTS FOR THE FIVE TRIALS

At the time of writing and using the information based on the five papers previously referenced,[9–13] the results in terms of vaccine efficacy are as given in Figure 2. In addition to the results quoted in the various papers, I have calculated confidence intervals by (a) using an exact argument based on a conditional binomial and the allocation ratio (two for AZ/Oxford and one in the other four cases) then (b) transferring the resulting interval for θ to one for VE by applying
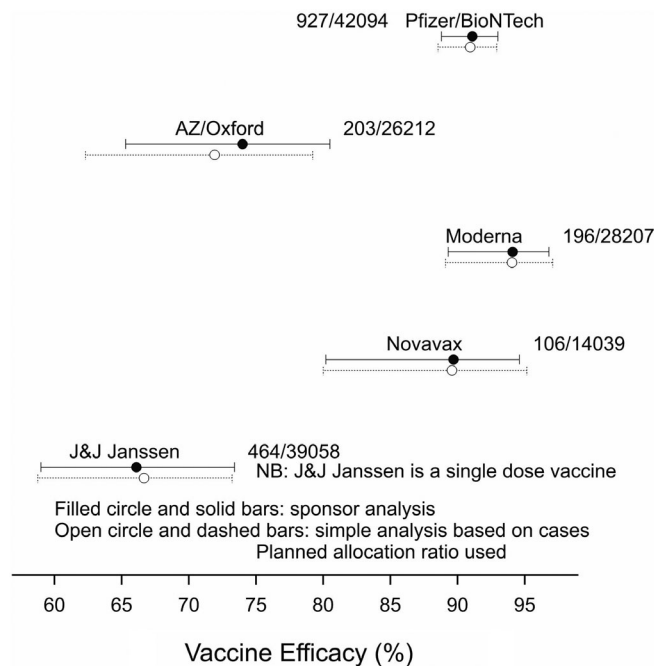

FIGURE 2 Results in terms of vaccine efficacy of the phase III trials listed in Table 1. For each of the five trials results of two analyses are shown. One is the published result and the other is a simple conditional analysis using the proportion of cases in the vaccine group, the panned allocation ratio and applying an exact binomial analysis. Also given are total numbers of cases and subjects.

TABLE 4 Numbers of subjects by trial and arm.

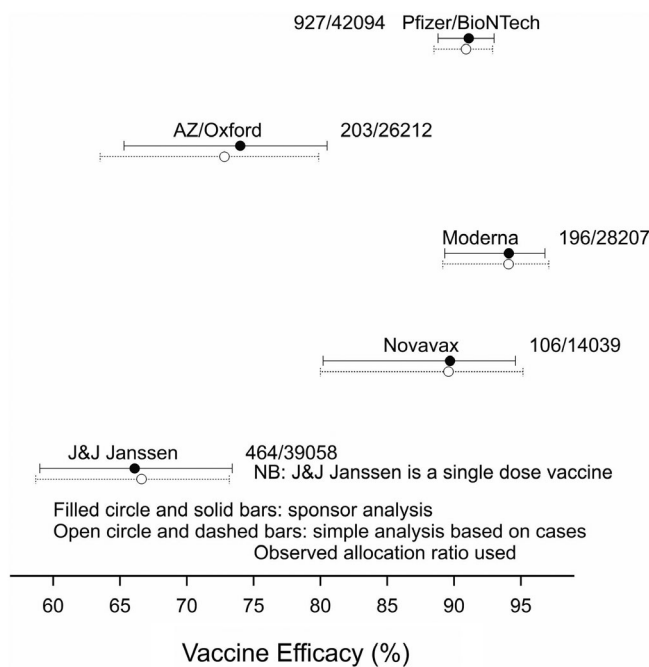| Sponsor	Planned Allocation Ratio	Control subjects n_c	Vaccine subjects n_v	Observed allocation ratio n_v/n_c	Chi-square	*p*-chi	*p*-exact
Pfizer/BioNTech	1	21,096	20,998	0.995	0.228	0.633	0.636
AZ/Oxford	2	8550	17,662	2.066	6.025	0.014	0.014
Moderna	1	14,073	14,134	1.004	0.132	0.716	0.721
Novavax	1	7019	7020	1.000	0.000	0.993	1.000
J&J Janssen	1	19,544	19,514	0.998	0.023	0.879	0.883

FIGURE 3 Representation of the results in Figure 2 but conditioning on the observed rather than the planned ratio of subjects.

Equation (7). The sponsors' estimates and confidence intervals are given by filled circles and solid bars and the result of my simple calculation by open circles and dashed bars. Totals numbers of cases and subjects are also indicated for the five studies. Although it is convenient to put these results in one table, the reader should be wary of assuming that they are comparable. Definitions of cases are not necessarily identical and time windows for observations differ and, in particular, as will be discussed below, follow-up has also differed and this can have an important effect on results.

Of course, one should prefer the estimates of the sponsor, which have the advantage of being pre-specified and using further (for example covariate or follow-up time) information. Nevertheless, it is interesting to see how close one can get using the simple argument. The one exception appears to be the AZ/Oxford trial. Part of the explanation may be that this is the trial for which the observed allocation ratio has the biggest difference from the expected one. The situation is shown in Table 4.

The chi-square statistic (with one degree of freedom) is what would apply as a test of the chance mechanism had simple randomisation been applied. In practice, some form of blocking will have been used by the sponsors and so the statistic may be regarded as an underestimate. Indeed, it is noticeable that the statistics are generally less than 1, so closer than one would expect had simple randomisation been applied but that there is one exception: the value for AZ/Oxford is six times what would be expected. Also given are the P-values calculated from the chi-square and the 'exact' p values calculated from a binomial test. These are in close agreement.

Of course, as I myself have argued previously,[23,24] using such baseline analyses is not an adequate way to decide whether changing the model by including a variable would change the analysis of the outcome variable. Furthermore, perhaps ironically, when such tests of balance are carried out, comparison of the sample size per arm is not the object but such values are, instead, conditioned on, say to compare if the distribution of covariates is 'balanced'. Nevertheless, the AZ/Oxford disparity is curious and, as far as I am aware, has not been explained by the authors. I find the discrepancy concerning, not because the P-value is less than 5% (it is very modest and one of five post-hoc values) but because I expect blocking to have been used in the trial, in which case, the allocation ratio ought to be closer to the expected ratio of 2:1.

The way to check if it has an effect, however, is to include it in the analysis. Figure 3 again gives the sponsor analysis but this time my alternative analysis conditions on the actual observed ratio of subjects. There is little difference to the situation in Figure 2, although my result for AZ/Oxford is now a little closer to the sponsor's. Of course, in making this adjustment, I am making a 'missing at random' assumption. Furthermore as already noted the denominator used by sponsors was often observations time and in any case, the numbers deemed to be at risk were sometimes a little

lower than the figures I have quoted. Nevertheless, in conclusion, what is remarkable is how close a simple analysis gets to the more sophisticated version used by the sponsors.

This raises an interesting issue. The exact analysis I have used does not rely on asymptotic results. To that extent, it might be considered to be superior to more flexible modelling that does. On the other hand, the more flexible modelling can take into account many relevant incidental matters that the exact analysis ignores. The supposed accuracy of the exact analysis is only achieved by disregarding these incidental matters in order to achieve a validity, which applies when averaged over all randomisations. (See Seven Myths of Randomisation for a criticism of this view.[25]) This too, therefore is a sort of asymptotic result, not as the sample size goes to infinity but as the randomisations do. As Jack Good put it, discussing what he called the Statistician's Stooge, "this precise objectivity is attainable, as always, only at the price of throwing away some information"[26] (p. 54). My intuition is that the modelling approach is to be preferred. I have presented the 'exact' analysis here not to propose it as a superior alternative but instead as a useful simple check.

5 | PFIZER/BIONTECH'S BAYESIAN APPROACH

Four sponsors nominated frequentist methods as their approach to analysing the planned trial. Pfizer/BioNTech, on the other hand, nominated a Bayesian approach. I now propose to discuss this in detail. Much of what I have to say concerns planning. However, since a plan is guided by the intended analysis, I shall also illustrate some points by analysing the results. Two points are worth noting. First, I do not have access to the covariate data that were used. Second, I shall illustrate the analysis using the outcome data that were available at the analysis published by Polack et al[27] in the New England Journal of Medicine on 10 December 2020. (As a referee has pointed out to me, although there have been subsequent analyses using more data, it is misleading to describe this as an interim analysis, since it was the analysis that was defined to take place after the planned target of 164 cases had accrued. Further follow-up since that time has provided more data but that is another matter).

As regards the first point, which in any case follows from necessity, I have already explained one can get very close to sponsor results by using the results from cases alone and this simple analysis has the virtue of allowing one to concentrate on what the essential difference between a frequentist and Bayesian analysis is. As regards the second, the more data one has, then, other things being equal, the smaller the difference between frequentist and Bayesian approaches. Using fewer data is thus more revealing. In any case, as discussed above they can be considered as those that were relevant to the primary analysis.

As the Polack et al paper[27] puts it "The 95.0% credible interval for vaccine efficacy and the probability of vaccine efficacy greater than 30% were calculated with the use of a Bayesian beta-binomial model." (P2605) Which prior distribution was used to produce the relevant posterior distribution is not explained at this point. However, a footnote to Table 2 states, "The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time." A further footnote to this table states, "Posterior probability was calculated with the use of a beta-binomial model with prior beta (0.700102, 1)." The quoted parameters are those for the prior beta distribution for a beta binomial analysis. It is this distribution I propose to discuss.

A striking feature of the distribution is that the first parameter is quoted to six significant figures and the second to only 1. To find the reason, it is necessary to consult the protocol,[28] which states:

"A minimally informative beta prior, beta (0.700102, 1), is proposed for $\theta = (1 - VE)/(2 - VE)$. The prior is centred at $\theta = 0.4118$ (VE = 30%) which can be considered pessimistic. The prior allows considerable uncertainty; the 95% interval for θ is (0.005, 0.964) and the corresponding 95% interval for VE is (−26.2, 0.995)" (pp. 102, 103).

A common form for parameterising the beta distribution is that given by Forbes et al[29] who write it as

$$f(X = x) = \frac{x^{\nu-1}(1-x)^{\omega-1}}{\int_0^1 u^{\nu-1}(1-u)^{\omega-1} du}, \quad 0 \le x \le 1 , \tag{8}$$
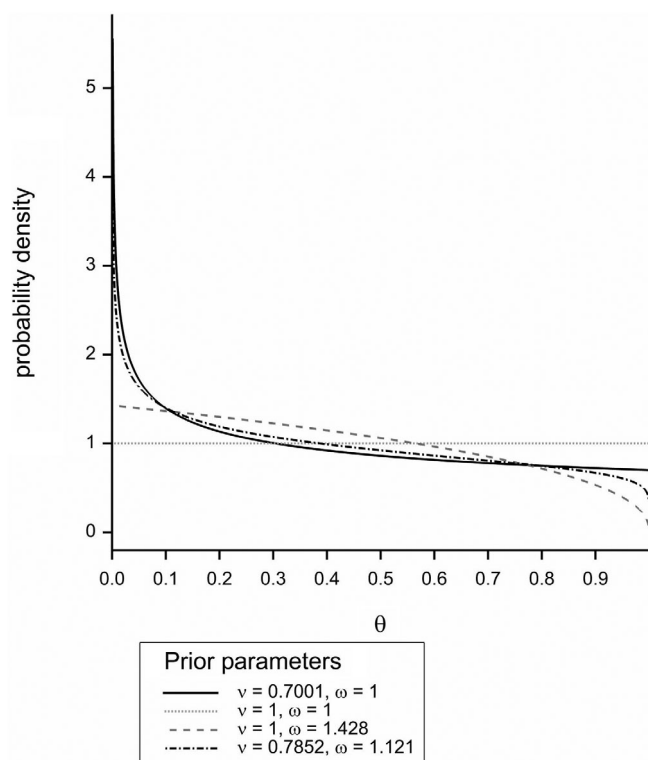
FIGURE 4 Prior distribution for θ used by Pfizer/BioNTech as well as various possible alternatives that might be considered.

where the denominator of (8) is the beta function. This appears to be the form that Pfizer/BioNTech have used. The argument x of the beta distribution ranges between 0 and 1, which makes it suitable to act as a quantity which is itself a probability. Note that vaccine efficacy is not a probability but the expected ratio of cases under vaccine to all cases *is* a probability and this is the parameter for which the prior distribution is established in the protocol and this is what we have labelled θ in (5). In, other words, we can substitute θ for x and $\nu = 0.700102$, $\omega = 1$ in (8). With this parameterisation, the mean of the prior distribution is 0.4118. The prior distribution in terms of θ is given by Figure 4. Also shown are various possible alternatives. We now proceed to discuss the choice of the prior distribution that was used.

Why the choice of this strange mean of 0.4118? The answer is that application of the transformation given by (7) yields a corresponding value of 30% on the vaccine efficacy scale and 30% was the minimum value of interest for vaccine efficacy that was commonly agreed (at least, by all five sponsors considered here). There are, however, three problems in giving this as a justification. The first is that it does not follow from the fact that the transformation of the mean of 0.4118 on the θ scale yields a value of 30% on the VE scale that the mean on the VE scale is 30%. For any monotonic transformation, the median on one scale can be transformed to the median on another but transformations over means are not necessarily the same as means over transformations.

The second problem is that there are infinitely many beta-distributions with a mean of 0.4118. Why was the one with $\nu = 0.700102$, $\omega = 1$ chosen? The choice of possible values is illustrated in Figure 5, which gives contours of the variance,

$$\text{var}(\nu, \omega) = \frac{\nu\omega}{(\nu + \omega)^2 (\nu + \omega + 1)} \qquad (9)$$

of the beta distribution as a function of $\nu \& \omega$. Note that the variance increases with lower values of the parameters that is to say towards the lower left hand side of the figure. The solid grey diagonal line rising from bottom left to top right joins all possible parameter combinations that have a mean of 0.4118. The diamond gives the combination chosen by Pfizer/BioNTech but many other choices are possible. A common default beta distribution is the one with $\nu = 1, \omega = 1$ for which the probability density is uniformly 1 for all values of θ and for which the variance is 1/12. Of course, this
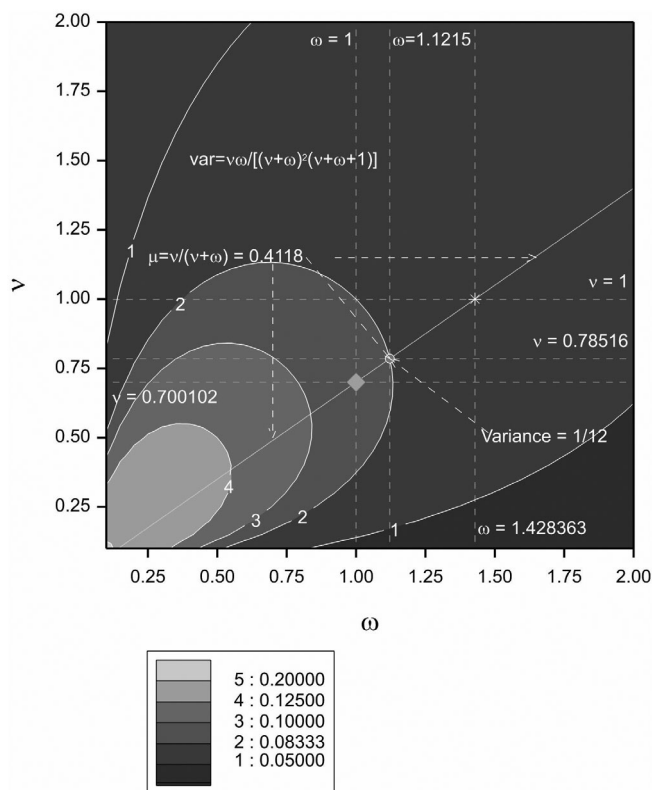
FIGURE 5 Contours of the variance of a beta distribution as a function of ν and ω. Contours 1–5 give various variances.

has a mean value of 0.5, not 0.4118. What Pfizer/BioNTech seem to have done is to change the value of ν and keep the value of ω the same in order to produce a mean of 0.4118. If they wanted to keep one of the parameters at 1 they could equally well have used $\nu = 1$, $\omega = 1.428363$, although this would have implied a smaller prior variance. This point is illustrated by an asterisk. However, if both values are changed in order to maintain the variance at 1/12 but yield a mean of 0.4116, the combination is $\nu = 0.78516$, $\omega = 1.1215$. This point is illustrated by a circle.

I am grateful to a referee for suggesting a justification. If the object is to target a particular quantile such as 0.4118 (which as previously noted corresponds to 30% vaccine efficacy) and one wishes to have a unimodal prior distribution and be minimally informative, then, as is explained in the appendix to Neuenschwander et al,[30] the following holds as regards a choice of ν and ω.

1. The quantile requirement must be satisfied.
2. At least one of the two parameters must be at least equal to one.
3. The sum of the parameters must be as low as possible.

The purpose of the first requirement is, of course obvious, the second ensures that the distribution is unimodal and the third minimises the information requirement, if this is defined by regarding the sum of the parameters as being a notional equivalent of the number of patients previously seen. As the referee points out, if this approach is adopted but the first requirement is made in terms of the mean rather than the median, then the solution is:

$$\nu = \frac{p}{1-p}, \omega = 1; \quad p \leq 0.5$$
$$\nu = 1, \omega = \frac{p}{1-p}; \quad p \geq 0.5,$$

(10)

where $p = E[X]$, and X is as defined in (8).

The third criticism is the most serious. There is no reason to choose 30% as a point to anchor the prior distribution. The value of 30% is a sort of agreed minimum relevant effect but it does not follow that it should represent prior belief. However, prior belief is what the prior distribution is supposed to represent. The value of 30% should be taken as a minimum target value for the posterior belief, for example by showing that the probability that vaccine efficacy is less than this can be no greater than 2.5%. This is a quite different matter from the prior belief and the statement in the protocol that this "can be considered pessimistic" is quite wrong. It can be considered a stringent standard to adopt for 'proof' of efficacy but a pessimistic prior distribution should surely not be centred on a vaccine having this efficacy. A more advanced Bayesian approach might use 30% to define some sort of loss function (although this would not be easy) but the issue of the prior distribution would be different from this.

The associate editor has reminded me of the important paper by Spiegelhalter, Freedman and Parmar[31] discussing Bayesian approaches to clinical trials. This explicitly distinguishes between plausible values, such as are relevant to prior belief, and those that are relevant to decision-making, such as zones of equivalence of treatments. It would be interesting to explore the implications of this but that would involve a considerable investigation, which I do not claim to be competent to undertake. In any case, in practical terms it does not matter. The distribution is extremely uninformative. Nevertheless, the definition of an uninformative distribution in terms of a parameter defined to six significant figures does seem somewhat strange.

6 | RESULTS OF THE BAYESIAN ANALYSIS

The reader is reminded that these are based on the results published in Polack et al.[27] at the planned termination of the trial and not the further results obtained by extensive follow-up. As already remarked, it is only necessary to know the number of cases under vaccine, Y_v, and under placebo, Y_p to perform an analysis that gets close to that which uses more information. At the time of this analysis, the values were $Y_v = 8$, $Y_p = 162$. If the prior distribution is a beta and the likelihood is binomial, then the posterior is also a beta distribution and to obtain its parameters one simply adds the respective numbers of cases to the respective parameters. (See, for example, Bernardo and Smith,[32] p. 271). Thus, we have

$$\nu_{\text{post}} = \nu_{\text{prior}} + Y_v = 0.700102 + 8 \approx 8.7,$$
$$\omega_{\text{post}} = \omega_{\text{prior}} + Y_p = 1 + 162 = 163. \tag{11}$$

The posterior mean of this distribution is given by[29]

$$\overline{\theta}_{\text{post}} = \frac{\nu_{\text{post}}}{\nu_{\text{post}} + \omega_{\text{post}}} = 0.0507 \tag{12}$$

and the posterior mode, say θ'_{post}, is given by[29]

$$\theta'_{\text{post}} = \frac{\nu_{\text{post}} - 1}{\nu_{\text{post}} + \omega_{\text{post}} - 2} = 0.0454. \tag{13}$$

The 2.5%, 97.5% and 50% points of the resulting posterior beta-distribution give the lower and upper limits of the so-called 95% credible interval and the median. The values I get for this are

$$\theta_{0.025} = 0.0232, \ \theta_{0.975} = 0.0880, \ \theta_{0.5} = 0.0489. \tag{14}$$

Of course, all these results are on the scale of the case proportion, θ, whereas what we want is the vaccine efficacy scale, VE. We can get results on this scale by applying (7) to (12), (13) and (14). Care must be taken with the latter since low values of θ correspond to high values of VE and vice versa. The posterior distribution is illustrated in Figure 6.

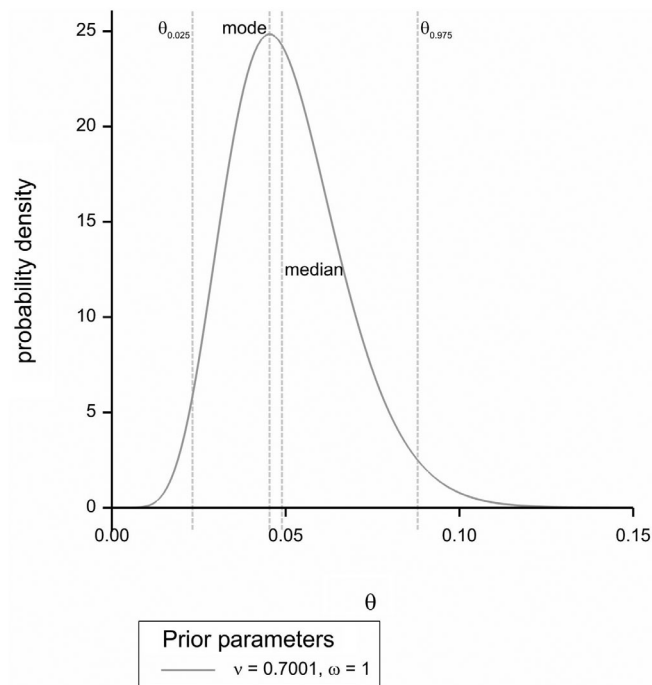However, we shall defer making the transformation until we have considered a possible frequentist analysis.

FIGURE 6 Posterior distribution of θ given the prior beta distribution used by Pfizer/BioNTech and the data analysed in December 2020.

7 | A FREQUENTIST ANALYSIS

We can use the binomial distribution given in (6) to calculate an 'exact' P-value for any observed number of cases under vaccine Y_v conditional on the total number of cases, $n = Y_v + Y_c$ in the vaccine and placebo group, for any assumed value θ. Here we have $Y_v = 8$ and $n = 8 + 162 = 170$, We simply vary θ to discover what values will give lower and upper tail areas of 0.025 at $Y_v = 8$. These are the upper and lower 95% confidence intervals.

Figure 7, which is taken from the second edition of *Dicing with Death*,[33] is a graphical representation as to how this works. The two curves plot the probability of obtaining 8 or fewer and 8 or more cases under vaccine as a function of the probability θ given that the total number of cases is 170. The dashed horizontal line at 0.025 indicates the desired tail area probability. The confidence limits are found from the points at which the two curves intersect this boundary and are indicated by vertical lines.

Using this approach, I obtain the following values for the lower and upper 95% confidence limits on the case proportion scale

$$\theta_{LCL} = 0.0205, \ \theta_{UCL} = 0.0906. \tag{15}$$

The point estimate is simply the case proportion, that is to say

$$\theta_{est} = \frac{Y_v}{n} = \frac{8}{170} = 0.04706 \tag{16}$$

and this is also illustrated in Figure 7.

8 | RESULTS ON THE VACCINE EFFICACY SCALE

We can translate all of these results to the vaccine efficacy scale by using the transformation given by (7), taking care to note that since vaccine efficacy increases as θ decreases we transform lower to upper confidence limits and vice versa.
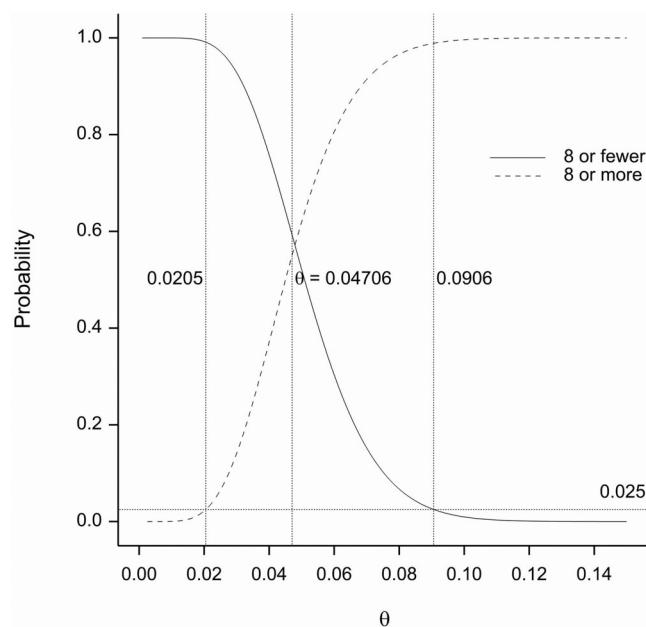
FIGURE 7 Confidence curves for the results of the Pfizer/BioNTech trial.

TABLE 5 Various Bayesian estimates and 95% credible intervals as well as a frequentist estimate and 95% confidence intervals.

Type	Point estimate (%)	Lower Limit (%)	Upper Limit (%)
Pfizer/BioNTech full model credible	95.0	90.3	97.6
Simple credible (mean θ)	94.7	90.4	97.6
Simple credible (mode θ)	95.2		
Simple credible (median θ)	94.9		
Simple confidence	95.0	90.0	97.9

Applying this to all the results we have so far we have them gathered in Table 5, which is taken from the second edition of *Dicing with Death*.[33]

All analyses except the first have been produced by me using just the fact that the cases on vaccine and placebo split 8–162. The first analysis is taken from the paper by Polack et al.[27] A curious feature about the reporting of this interval is that although the method to produce it is described as Bayesian and although Table 2 of that paper, in which it is produced, has the footnote, "The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time", elsewhere in the paper a *confidence interval* is referred to. The discussion of efficacy has the following statement

> "Among 36,523 participants who had no evidence of existing or prior SARS-CoV-2 infection, eight cases of Covid-19 with onset at least 7 days after the second dose were observed among vaccine recipients and 162 among placebo recipients. This case split corresponds to 95.0% vaccine efficacy (95% confidence interval [CI], 90.3 to 97.6."[27] (p. 2610, my emphasis).

This is presumably a typographical error and, of course, as Table 5 shows, in practice it makes little difference which method is used. Note also that other analyses in the paper apart from the main analysis are frequentist and planned to be so.

However, one point should be understood about the intervals presented, whether confidence or credible. They are statements about what happened in the trial not about what will happen in future. Confusion of these two very different matters is very common in the discussion of clinical trials[34] partly, I think, because it is incorrectly supposed that saying what happened is a matter of descriptive statistics only. However, if we wish to know whether what was

observed to happen was *caused* by the treatments given, then we are faced with the difficulty that we do not know what would have happened had subjects who were given the vaccine been given placebo and vice versa. It is this causal effect in the trial that the analyses attempt to address.

What will happen in future, is, of course contingent on many possible developments, as the history of the emergence of the delta variant and then the omicron variant has shown. Furthermore, the follow-up time of the trials is limited, so it is quite possible that efficacy will wane to a degree that makes the estimates obtained scarcely relevant, a point that is of great practical importance.

In any case, the vaccines are now being used in ways that have not been studied formally. To strike a personal note, at the time of writing, I have received two doses of the AZ/Oxford vaccine, in the form studied in the large trial described here and a follow up 'booster' low dose of the Moderna vaccine. I am happy to have done so but, as far as I am aware, this combination had not been formally studied at the time I was offered it. This raises the issue that one should be careful about comparing the efficacy of the J&J Janssen vaccine to the others. Figure 2 suggests that it had lower vaccine efficacy. However, it was given as a single dose, whereas the others used two doses. It may be that the J&J Janssen vaccine would have similar efficacy to the others if given in two doses. One might argue that this is not for discussion since it has not been studied let alone proven. However, if we are prepared to judge, without formal demonstration, that adding a Moderna booster to a double dose of the AZ/Oxford vaccine is beneficial, can this be a reasonable objection? What would be useful would be to have information on the protective value of the other vaccines after a single dose but, the time window between first and second dose being relatively short, the number of control group cases that arise within the interval may be inadequate to allow a reasonable calculation and in any case, given that efficacy may wane eventually with time, the short window may be misleading.

The paper by Thomas et al.[9] presents an estimate (confidence limits) of efficacy for the Pfizer/BioNTech vaccine in the interval between first and second dose of only 58.4% (40.8%–71.2%). However, if only cases after the first 11 days post vaccination are considered, this rises to 91.7% (79.6%–97.4%).[9] This suggests that the Pfizer/BioNTech vaccine might be very effective in a single dose but is far from conclusive.

The practical point is that if we are willing to give subjects three vaccine shots when we have only studied two, we must presumably be prepared to give subjects two shots although we have only studied one. That being so one could argue that perhaps a more efficient approach would have been for every sponsor to have adopted the J & J Janssen single dose approach for studying efficacy, even if the option to use two doses was considered important. This might have made trials logistically simpler and consequently quicker to conclude. Indeed, as the pandemic unrolled, some countries adopted the policy of trying to vaccinate as many subjects as possible once, even if that meant that the delay between first and second doses was longer for some subjects than that which had been observed in the trials.

In short, we need much more than just clinical trial results to make practical decisions. Cynics may claim that such considerations render clinical trials pointless but this, I consider, is unreasonable. We have no choice but to make decisions about treatment and although clinical trials provide no guarantees, they do eliminate many biases that would otherwise make such choices even more uncertain.

9 | DISCUSSION AND LESSONS

I consider that the response to the COVID-19 pandemic by the pharmaceutical industry and to some degree academia (bearing in mind the AstraZeneca and Oxford University collaboration) has been very impressive. Statisticians should not delude themselves that statistics is the major part of this story. It is the work of the life scientists that deserves particular praise. Nevertheless, running the clinical trials to a successful conclusion (by which I mean a conclusion that is informative, whether the news is good or bad) does require careful planning and statistics makes a large contribution to that planning.

The trials illustrated the value of concurrent control. Infection rates were changing throughout the period in which the trials were being run and in any case could be expected to vary from place to place. Infection rates in the control groups varied markedly from trial to trial, as is illustrated by Figure 8. The higher rate in the Pfizer/BioNTech compared to the others is quite marked but then the follow-up at the time of reporting is different and this can have very important consequences[15] not only in terms of changing susceptibility over time, for example due to new variants or the introduction or relaxation of various other measures, but because of the complex relationship between time to event variables and binary dichotomies. Also shown are fixed and random effects meta-analyses over the five trials. The considerable extra-binomial variation (heterogeneity) results in the confidence interval for the random effects estimate
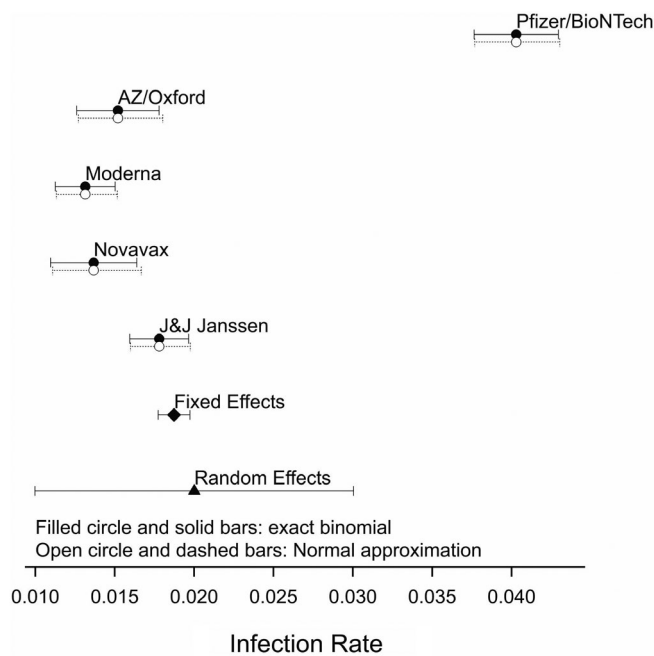
FIGURE 8 Infection rates in the placebo groups with associated confidence intervals for the five trials. Closed circles and solid bars, Normal approximation. Open circles and dashed bars, exact binomial. The fixed effects meta-analysis of all five trials is indicated by a diamond and the random effects meta-analysis by a triangle.

being much wider than those for the fixed effects estimate. This does not mean that either estimate is particularly useful but it points to the degree of variation.

As shown in Table 2, three of the trials, those of Pfizer/BioNTech, Moderna and Novavax, were described as observer blind with the other two, those of AZ/Oxford and J&J Janssen being described as double blind. Blinding is important. It not only serves to avoid biased judgement of outcomes but helps to support the randomisation process. In a blog, I wrote on the subject[35] I gave the following list of possible issues:

1. Subjects who are chosen to be vaccinated are invited to attend a clinic to receive their vaccine.
2. A team of health workers is assigned for vaccination and another team is assigned for assessing controls.
3. Blood samples are collected by the vaccinating clinic.
4. Subsequent blood samples (say after 28 days) are also collected in the vaccinating clinic.
5. Samples are sent in batches to the laboratory to be analysed.
6. Control subjects are visited by nurses at home to collect blood samples.

And discussed these as follows:

> "All of these have the ability to subvert the randomisation process. With the exception of the first and the last and possibly the second, they are not biasing *per se*. But they make very debatable any assumption of independence that might be naively made in analysis. A biasing factor is one that prevents the estimate from converging on the 'right' answer as the sample size grows but attracts it to some other value. Lack of independence, on the other hand, affects the rate of convergence."

Such problems cannot arise if the trials are double blind, since whatever nuisance effect may currently apply to a group of subjects it is impossible for them to be clustered except accidentally, for which randomisation allows, since they have been randomly allocated to one group or another and since blinding prevents identification of subjects in a way that could affect subsequent handling. Whether observer blind trials provide adequate protection against such problems is not clear to me.

On the whole, I think the plans here were good. Of course, one must give credit to the extremely successful logistic operations that were mounted in the face of many difficulties. Many resources had to be combined energetically and
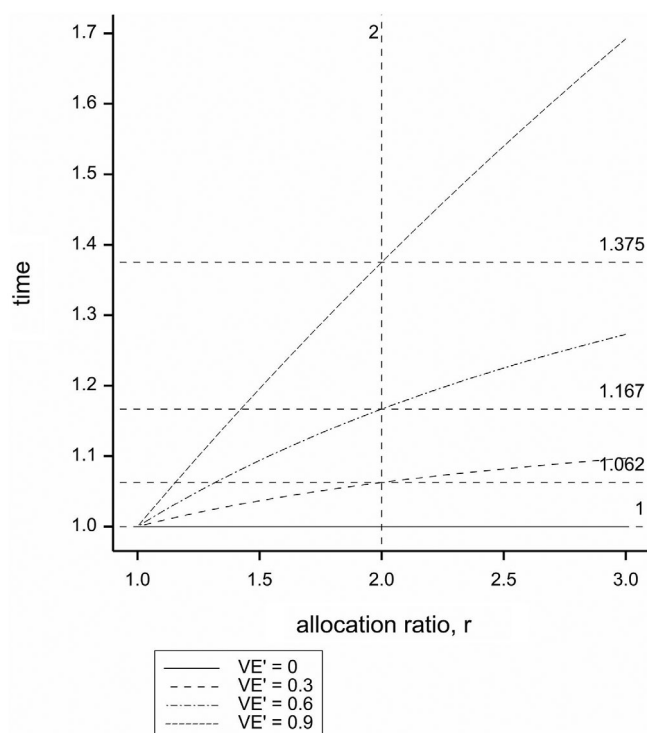
FIGURE 9   Time to recruit a target number of cases as a function of the allocation ratio (vaccine subjects to control subjects) for various degrees of vaccine efficacy as a proportion (not a percentage). Time is expressed as a ratio of the time it would take for a 1:1 allocation.

effectively. Statisticians have made contributions to planning multi-centre trials[36,37] but running them well is very much more than just statistics. Nevertheless, sample size determination and other aspects of design were important.

In fact, the numbers of subjects required was secondary. What was needed to get adequate power and precision were cases. Subject numbers were targeted to deliver the required number of cases using assumptions of background rates. In my opinion, it was an unwise decision of AZ/Oxford to allocate two subjects to vaccine for every one allocated to placebo. If $VE' = VE/100$ is vaccine efficacy expressed as a proportion rather than a percentage and R is the ratio of subjects allocated to vaccine rather than placebo, and $time_R$ is the ratio of the time it will take to reach a target number of cases compared to what the time would be for an allocation ratio of $R = 1$, then

$$Time_R = \frac{(2 - VE')(R + 1)}{2(R - R(VE') + 1)}. \tag{17}$$

This is equal to 1 (trivially) if $R = 1$ but also if $VE' = 0$. However, if $VE' > 0$, it is an increasing function of R. The situation is illustrated in Figure 9, which plots expression (17) for various values of vaccine efficacy. Shown are the cases where the vaccine is inefficacious, vaccine efficacy is 30% as assume for the 'null hypothesis, or 60% as was assumed for the power calculation or 90%, the sort of value observed in some trials by some sponsors. I estimate that for the target efficacy of 60% assumed in the AZ/Oxford protocol, the 2:1 allocation would prolong the duration of the trial by 1/6, other things being equal.

Where binary data are concerned because the precision is not independent of the parameters, there can be a justification for unequal allocation ratios. However, I have my doubts that this can be a justification here. For discussion of allocation ratios for vaccine trials see papers by Patterson et al.[15,38]

The Pfizer/BioNTech trial illustrates a number of matters of theoretical interest that turn out, however, to be unimportant practically. The first is that the choice of prior distribution may be difficult. The second is that there is the danger of confusing prior belief and desired objectives, for example to prove that efficacy is at least 30%. The third is that it can be difficult to translate belief from one scale to another, especially if the 'anchoring' is done in terms of moments. Models over means are not the same as means over models. In making this criticism, I am not belittling the

work of the statisticians in the Pfizer/BioNTech trial. They helped to produce a fine protocol and delivered an excellent analysis. I am extremely impressed by the speed with which an effective plan was delivered.

Of course, in the end it did not matter. Even for the earlier analysis based on 8 and 162 cases that I chose to present here, the prior distribution was overwhelmed by the likelihood. For the results available towards the end of 2021, 77 cases under vaccine and 850 under placebo, the influence of the prior distribution was even less.

I also think one further very important matter was illustrated by the vaccine trials. Statistical work is most impressive when it is practically grounded. Andy Grieve has never shied away from theoretical difficulties when they have presented themselves. On the other hand, he has also always been motivated by practical problems, as the paper with Amy Racine, Hugo Flühler and Adrian Smith that I cited earlier shows. The COVID-19 pandemic has presented statisticians with many challenges. Theory guided by application has been the way to meet them. I can only applaud the response of the pharmaceutical statistics profession to the challenge.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analysed in this study. Copies of the Genstat programs used to produce the figures are obtainable from me on request.

ORCID
Stephen Senn https://orcid.org/0000-0002-7558-8473

REFERENCES
1. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc*. 1990;85(410):398-409.
2. Racine A, Grieve AP, Fluhler H, Smith AFM. Bayesian methods in practice—experiences in the pharmaceutical-industry. *App Stat: J R Stat Soc Ser C*. 1986;35(2):93-150.
3. Grieve AP. The two-period changeover design in clinical trials. *Biometrics*. 1982;38(2):517.
4. Grizzle JE. The two-period change over design and its use in clinical trials. *Biometrics*. 1965;21:467-480.
5. Grieve AP. Tests of sphericity of normal-distributions and the analysis of repeated measures designs. *Psychometrika*. 1984;49(2):257-267.
6. Grieve AP. Pretest-posttest designs. *Am Stat*. 1981;35(3):177-178.
7. Faulds D, Hollingshead LM, Goa KL. Formoterol: a review of its pharmacological properties and therapeutic potential in reversible obstructive airways disease. *Drugs*. 1991;42:115-137.
8. Grieve A, Senn S. Estimating treatment effects in clinical crossover trials. *J Biopharm Stat*. 1998;8(2):191-233. discussion 235-47.
9. Thomas SJ, Moreira ED Jr, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine through 6 months. *N Engl J Med*. 2021;385:1761-1773. doi:10.1056/NEJMoa2110345
10. Falsey AR, Sobieszczyk ME, Hirsch I, et al. Phase 3 safety and efficacy of AZD1222 (ChAdOx1 nCoV-19) Covid-19 vaccine. *N Engl J Med*. 2021;385:2548-2360. doi:10.1056/NEJMoa2105290
11. Baden LR, El Sahly HM, Essink B, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med*. 2021;384(5):403-416. doi:10.1056/NEJMoa2035389
12. Heath PT, Galiza EP, Baxter DN, et al. Safety and efficacy of NVX-CoV2373 Covid-19 vaccine. *N Engl J Med*. 2021;385(13):1172-1183. doi:10.1056/NEJMoa2107659
13. Sadoff J, Gray G, Vandebosch A, et al. Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *N Engl J Med*. 2021; 384(23):2187-2201. doi:10.1056/NEJMoa2101544
14. Huitfeldt A, Fox MP, Daniel RM, Hróbjartsson A, Murray EJ. Shall we count the living or the dead? ArXiv Preprint arXiv:210606316v3 2021. Accessed 8 March 2022. https://arxiv.org/abs/2106.06316
15. Patterson S, Fu B, Meng Y, Bailleux F, Chen J. Statistical observations on vaccine clinical development for pandemic diseases. *Stat Biopharm Res*. 2022;14(1):28-32. doi:10.1080/19466315.2021.1919197
16. Dragalin V, Fedorov V. Multistage designs for vaccine safety studies. *J Biopharm Stat*. 2006;16(4):539-553.
17. Dragalin V, Fedorov V, Cheuvart B. Statistical approaches to establishing vaccine safety. *Stat Med*. 2002;21(6):877-893. doi:10.1002/sim.1039
18. Pocock SJ. Group sequential methods in design and analysis of clinical trials. *Biometrika*. 1977;64(2):191-209.
19. Lan KKG, Demets DL. Discrete sequential boundaries for clinical-trials. *Biometrika*. 1983;70(3):659-663.
20. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549-556.
21. Lehmann EL. *Testing Statistical Hypotheses*. 2nd ed. Chapman and Hall; 1994.
22. McCullagh P, Nelder JA. *Generalized Linear Models*. Vol 511. 2nd ed. Chapman and Hall; 1989.

23. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Stat Med*. 1989;8(4):467-475.
24. Senn SJ. Testing for baseline balance in clinical trials. *Stat Med*. 1994;13(17):1715-1726.
25. Senn SJ. Seven myths of randomisation in clinical trials. *Stat Med*. 2013;32(9):1439-1450. doi:10.1002/sim.5713
26. Good IJ. *Good Thinking: the Foundations of Probability and its Applications*. University of Minnesota Press; 1983.
27. Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med*. 2020;383(27):2603-2615. doi:10.1056/NEJMoa2034577
28. BioNTech/Pfizer. A phase 1/2/3, placebo-controlled, randomized, observer-blind, dose-finding study to evaluate the safety, tolerability, immunogenicity, and efficacy of SARS-Cov-2 RNA vaccine candidates against Covid-19 in healthy individuals. 2020:1–136. Accessed 28 November 2021. https://cdn.pfizer.com/pfizercom/2020-11/C4591001_Clinical_Protocol_Nov2020.pdf
29. Forbes C, Evans M, Hastings N, Peacock B. *Statistical Distributions*. John Wiley & Sons; 2011.
30. Neuenschwander B, Branson M, Gsponer T. Critical aspects of the Bayesian approach to phase I cancer trials. *Stat Med*. 2008;27(13):2420-2439. doi:10.1002/sim.3230
31. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials. *J R Stat Soc Ser A: Stat Soc*. 1994;157:357-387.
32. Bernardo JM, Smith AF. *Bayesian theory*. Vol 405. 1st ed. John Wiley & Sons; 1994:586.
33. Senn SJ. *Dicing with Death*. 2nd ed. Cambridge University Press; 2022.
34. Senn SJ. Added values: controversies concerning randomization and additivity in clinical trials. *Stat Med*. 2004;23(24):3729-3753.
35. Senn S. Blind date. 2021, Accessed 17 January 2021. https://www.linkedin.com/pulse/blind-date-stephen-senn/
36. Anisimov VV, Fedorov VV. Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. *Stat Med*. 2007;26(27):4958-4975. doi:10.1002/sim.2956
37. Anisimov VV, Fedorov VV. Design of multicentre clinical trials with random enrolment. In: Auget JL, Balakrishnan N, Mesbah M, Molenberghs G, eds. *Advances in Statistical Methods for the Health Sciences: Applications to Cancer and Aids Studies, Genome Sequence Analysis, and Survival Analysis*. Birkhauser; 2007:387-400.
38. Patterson S, Bailleux F, Chen J, Zhu M. Imbalanced randomization in vaccine clinical safety trials. *Stat Biopharm Res*. 2020;12(3):279-283.