Clinical Infectious Diseases

INVITED ARTICLE

IDSA Infectious Diseases Society of America | hivma hiv medicine association | OXFORD

INNOVATIONS IN DESIGN, EDUCATION ANDANALYSIS (IDEA): Victor De Gruttola and Scott R. Evans, Section Editors

Evaluating the Efficacy of Coronavirus Disease 2019 Vaccines

Dan-Yu Lin,[1] Donglin Zeng,[1] Devan V. Mehrotra,[2] Lawrence Corey,[3] and Peter B. Gilbert[3]

[1]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA; [2]Biostatistics & Research Decision Sciences, Merck & Co, Inc, North Wales, Pennsylvania, USA; and [3]Vaccine and Infectious Disease Division, Fred Hutch, Seattle, Washington, USA

A large number of studies are being conducted to evaluate the efficacy and safety of candidate vaccines against coronavirus disease 2019 (COVID-19). Most phase 3 trials have adopted virologically confirmed symptomatic COVID-19 as the primary efficacy end point, although laboratory-confirmed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is also of interest. In addition, it is important to evaluate the effect of vaccination on disease severity. To provide a full picture of vaccine efficacy and make efficient use of available data, we propose using SARS-CoV-2 infection, symptomatic COVID-19, and severe COVID-19 as dual or triple primary end points. We demonstrate the advantages of this strategy through realistic simulation studies. Finally, we show how this approach can provide rigorous interim monitoring of the trials and efficient assessment of the durability of vaccine efficacy.

Keywords. multiple primary end points; phase 3 trials; SARS-CoV-2; severe COVID-19; totality of evidence.

There is an urgent need to develop effective vaccines against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus causing the global coronavirus disease 2019 (COVID-19) pandemic. Several candidate vaccines have shown strong immune responses and acceptable safety profiles and have moved rapidly into large-scale phase 3 trials [1–8]. As of 8 December 2020, 28 phase 3 trials on 13 candidate vaccines had been launched around the world [7]. Through Operation Warp Speed, the US government selected several of these candidates for phase 3 testing, including mRNA vaccines (mRNA-1273, BNT162b1) that encode the prefusion stabilized SARS-CoV-2 Spike protein [2, 3], a recombinant replication-defective chimpanzee adenovirus that expresses a wild-type SARS-CoV-2 Spike protein (AZD1222) [4], a recombinant, replication-incompetent adenovirus type 26 (Ad26) vector vaccine that encodes a stabilized SARS-CoV-2 Spike protein (Ad26.COV2.S) [5], a SARS-CoV-2 recombinant stabilized Spike protein vaccine with AS03 adjuvant, and a SARS-CoV-2 recombinant stabilized Spike protein nanoparticle vaccine (SARS-CoV-2 rS) with Matrix-M1 adjuvant [6].

The vaccine regimens have generally protected against COVID-19 end points in animal models [5] and have induced binding and neutralizing antibody responses to vaccine-insert Spike proteins in most vaccine recipients, exceeding response levels seen in convalescent sera [2–4, 6]. The antibody marker end points are of the types that have been accepted as surrogate end points for many approved vaccines [9], generating enthusiasm that the vaccines can plausibly confer protection. Interim results from Pfizer/BioNTech, Moderna, and AstraZeneca/Oxford University suggested high vaccine efficacy against COVID-19.

Rapid introduction of effective vaccines in the United States and other countries with high numbers of COVID-19 cases would be a major step toward halting the global pandemic. However, deployment of a noneffective vaccine could actually worsen the pandemic because public acceptance of a COVID-19 vaccine might diminish the implementation of other control measures. Thus, we need speedy and reliable evaluation of the efficacy of COVID-19 vaccines on the basis of clinically relevant end points.

Most phase 3 trials have adopted virologically confirmed symptomatic COVID-19 illness as the primary efficacy end point, although laboratory-confirmed SARS-CoV-2 is also acceptable [10]. It is possible that a vaccine is much more effective in preventing severe than mild COVID-19. Thus, we should also evaluate the effect of vaccination on severe COVID-19 [10]. However, a large sample size is likely required for a trial that uses a severe COVID-19 end point.

We propose using SARS-CoV-2 infection, symptomatic COVID-19, and severe COVID-19 as triple primary end points or using SARS-CoV-2 infection and symptomatic COVID-19 or symptomatic COVID-19 and severe COVID-19 as dual primary end points, the specific choice depending on the expected incidence of the 3 events and on the targeted vaccine efficacy for the 3 end points. This approach incorporates more evidence on vaccine efficacy into decision making than using only 1 of the 3 events as the primary end point. It can improve statistical power and increase the likelihood of meeting vaccine success criteria, thus accelerating the discovery and licensure of effective vaccines.

Received 23 September 2020; editorial decision 10 December 2020; published online 18 December 2020.

Correspondence: Dan-Yu Lin, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420 (lin@bios.unc.edu).

Clinical Infectious Diseases® 2021;73(8):1540–4

DOI: 10.1093/cid/ciaa1863

METHODS

We consider the end points of SARS-CoV-2 infection, symptomatic COVID-19, and severe COVID-19, referring to them as infection, disease, and severe disease, respectively. Suppose that a large number of individuals are randomly assigned to vaccine or placebo and that the trial records whether or not each participant has developed each of the 3 end points by the end of follow-up, as well as their length of follow-up.

We formulate the effect of the vaccine on each of the 3 end points through a Poisson model. Although investigators are mainly interested in the first occurrence of each event, the Poisson modeling approach provides a reasonable approximation to the data because the event rates for all 3 end points are relatively low. We define the vaccine efficacy in terms of the proportionate reduction in the event rate between vaccinated and unvaccinated individuals.

The criteria for claiming that a vaccine is successful should be strict enough to ensure worthwhile efficacy. A vaccine with an efficacy that is higher than 50% can markedly reduce the incidence of COVID-19 among vaccinated individuals and help to build herd immunity. An advisory panel convened by the World Health Organization (WHO) recommended 50% vaccine efficacy for at least 6 months post-vaccination as a minimal criterion to define an efficacious vaccine [11]. The US Food and Drug Administration (FDA) guidance defines vaccine success criteria as a point estimate of vaccine efficacy at least 50% and the interim-monitoring adjusted lower bound of the 95% confidence interval exceeding 30% [10]. The FDA guidance criteria do not specify a minimum period of follow-up. However, given the intent of current vaccine development to identify efficacious vaccines within several months of trial initiation, the expectation seems to be reliable evidence for vaccine efficacy over approximately 6 months, consistent with the WHO recommendation.

Many phase 3 trials specify assessment of vaccine efficacy over longer-term follow-up as an important study objective. The FDA guidance document states that "a lower bound ≤30% but >0% may be acceptable as a statistical success criterion for a secondary efficacy endpoint, provided that secondary endpoint hypothesis testing is dependent on the success on the primary endpoint." This statement refers to earlier FDA guidance on a fixed-sequence testing method [12], under which vaccine efficacy is tested against a sequence of secondary end points in a predefined order where tests of each end point are performed at the same significance level (1-sided type I error of 2.5%), moving to the next end point only after a success on the previous end point. The WHO Solidarity Trial protocol [13] specifies symptomatic COVID-19 through longer-term follow-up (ideally 12 months or longer) and severe COVID-19 over the same time frame as secondary end points. Following these guidelines and precedents, we consider hypothesis testing of vaccine efficacy over 12 months as a secondary analysis, using a null hypothesis that is less stringent than the 30% null hypothesis value used for the primary analysis, recognizing that it is more difficult for a vaccine to provide 12-month than 6-month protection and that even moderate vaccine efficacy through 12 months could be an important characteristic of a COVID-19 vaccine. In sum, we consider both the assessment of vaccine efficacy against primary end points over 6 months using a 30% null hypothesis and the assessment of vaccine efficacy against the same end points over 12 months using a 0% or 15% null hypothesis.

For each of the 3 end points, we obtain the maximum likelihood estimator for the vaccine efficacy under the Poisson model. In addition, we calculate the score statistic for testing the null hypothesis that the vaccine efficacy is less than a certain lower limit, say 30%, against the alternative hypothesis that the vaccine efficacy is greater than the lower limit. We divide the score statistic by its standard error to create a standard-normal test statistic.

We propose to test all 3 null hypotheses, adjusting the significance threshold for the 3 test statistics to control the overall type I error at the desired level. We consider a vaccine to be successful if any of the 3 null hypotheses is rejected. We describe this multiple testing method in greater detail in Supplementary Appendix 1, where we also describe a sequential testing procedure to determine which of the 3 null hypotheses should be rejected.

In the sequential testing procedure, we order the 3 hypotheses according to the order of the 3 observed test statistics, from the most extreme observed value to the least extreme. We test the first null hypothesis using the significance threshold from the aforementioned multiple testing procedure. If the first null hypothesis is rejected, we test the second null hypothesis by applying the multiple testing procedure to the remaining 2 test statistics. If the second null hypothesis is rejected, we test the last null hypothesis by using the unadjusted significance threshold.

Clearly, this sequential testing procedure is more powerful than the multiple testing procedure in identifying which end points the vaccine is efficacious against. Both the proposed multiple testing and sequential testing methods properly account for the correlations of the test statistics and thus are more powerful than the conventional Bonferroni correction and related multiplicity adjustments that assume independence of tests.

If the effects of a vaccine are expected to be similar among the 3 end points, then we can enhance statistical power by combining the evidence of the vaccine effects on the 3 end points and performing a single test of overall vaccine efficacy. Specifically, we propose taking the sum of the 3 score statistics and dividing the sum by its standard error to create a standard-normal test statistic. We refer to this method as the combined test (Supplementary Appendix 1); this is in the same vein as combining estimators for a common effect in meta-analysis [14].

Instead of the triple primary end points, we may consider the dual primary end points of infection and disease if severe

disease is very rare or the dual primary end points of disease and severe disease if the vaccine is expected to be only weakly effective against infection. Clearly, the above methods can be modified to test only 2 of the 3 end points.

It is desirable to periodically examine the accumulating data from a phase 3 trial, so that the trial can be terminated if sufficient evidence emerges for a highly effective vaccine or a weakly effective candidate. In order to obtain rigorous stopping boundaries for a trial, we need to derive the joint distribution of the test statistics over interim looks. In Supplementary Appendix 2, we show that the proposed test statistics over interim looks are jointly normal with the independent increment structure such that standard methods for interim analyses [15–18] can be applied.

RESULTS

First, we conducted a series of simulation studies to compare the performance of the proposed methods with the use of a single primary end point in evaluating short-term vaccine efficacy. We assigned 27 000 participants to vaccine or placebo at a ratio of 1:1. We assumed that participants were enrolled at a constant rate over a 2-month period and that vaccine efficacy was evaluated 6 months after the first participant was enrolled. We let 1% of the placebo participants acquire infection, 0.6% experience disease, and 0.12% develop severe disease (Supplementary Appendix 3). These event proportions were based on the assumption of annualized incidence of about 1.5% for symptomatic COVID-19 in the placebo group, together with the assumptions that about 40% of infections are asymptomatic and that about 20% of symptomatic COVID-19 cases will be severe. We set the vaccine efficacy for disease, denoted by VE_D, to 60%; we set the vaccine efficacy for infection, denoted by VE_I, to 40%, 50%, 55%, or 60%; and we set the vaccine efficacy for severe disease, denoted by VE_S, to 60%, 70%, 80%, or 90% (Supplementary Appendix 3). For each combination of VE_I, VE_D, and VE_S, we simulated 100 000 datasets. (The average number of each end point can be easily calculated. For example, there are approximately 189 cases of infection, 113 cases of disease, and 23 cases of severe disease under $VE_I = VE_D = VE_S = 0.6$.) In each dataset, we tested the null hypothesis that the vaccine efficacy is at most 30% against the alternative hypothesis that the vaccine efficacy is greater than 30% at the 1-sided nominal significance level of 2.5%.

Table 1 summarizes the power of various methods for testing the null hypothesis of no worthwhile efficacy (ie, at most 30%). Use of the single end point of disease has 80% power under $VE_D = 60\%$. Indeed, we chose the sample size and disease rate in the placebo group to achieve this power, which is considered the benchmark for other methods. When VE_I is equal to or slightly below VE_D, the single end point of infection is more powerful than the single end point of disease (eg, 96% vs 80% power under $VE_I = VE_D = 60\%$) because infection is more frequent than disease. Due to low incidence, the single end point of severe disease has poor power unless VE_S is very high (eg, 69% and 91% power under $VE_S = 80\%$ and 90%, respectively). The combined test for the dual end points of infection and disease and the combined test for the triple end points are substantially more powerful than using disease as the single end point when VE_I is similar to VE_D (eg, 94% and 93% power for the 2 combined tests vs 80% power for the single end point of

Table 1. Statistical Power (%) for Testing the Null Hypothesis of At Most 30% Vaccine Efficacy Against Infection, Disease, and Severe Disease Over 6 Months

Vaccine Efficacy			Single End Point			Combined Test			Multiple Testing			Bonferroni		
VE_I	VE_D	VE_S	I	D	S	I-D	D-S	I-D-S	I-D	D-S	I-D-S	I-D	D-S	I-D-S
40%	60%	60%	21	80	27	51	77	53	75	75	72	73	74	69
40%	60%	70%	21	80	45	51	83	57	75	78	75	73	77	72
40%	60%	80%	21	80	69	51	88	61	75	85	82	73	84	79
40%	60%	90%	21	80	91	51	93	65	75	94	92	73	93	90
50%	60%	60%	65	80	27	78	77	78	79	75	76	77	74	73
50%	60%	70%	65	80	45	78	83	81	79	78	78	77	77	75
50%	60%	80%	65	80	69	78	88	83	79	85	84	77	84	81
50%	60%	90%	65	80	91	78	93	86	79	94	93	77	93	91
55%	60%	60%	84	80	27	87	77	86	86	75	83	84	74	80
55%	60%	70%	84	80	45	87	83	89	86	78	84	84	77	82
55%	60%	80%	84	80	69	87	88	91	86	85	88	84	84	87
55%	60%	90%	84	80	91	87	93	93	86	94	95	84	93	94
60%	60%	60%	96	80	27	94	77	93	94	75	92	93	74	91
60%	60%	70%	96	80	45	94	83	94	94	78	93	93	77	92
60%	60%	80%	96	80	69	94	88	96	94	85	95	93	84	94
60%	60%	90%	96	80	91	94	93	97	94	94	98	93	93	97

VE_I, VE_D, and VE_S denote, respectively, the vaccine efficacy for infection, disease, and severe disease. I, D, and S denote, respectively, infection, disease, and severe disease. I-D, D-S, and I-D-S denote, respectively, the dual end points of infection and disease, the dual end points of disease and severe disease, and the triple end points of infection, disease, and severe disease. The power pertains to a single test at the 1-sided nominal significance level of 2.5%.

disease under $VE_I = VE_D = VE_S = 60\%$). The combined test for the dual end points of disease and severe disease is more powerful than the single end point of disease when VE_S is high (eg, 93% vs 80% power under $VE_S = 90\%$). The combined test is more powerful than multiple testing for the dual end points of disease and severe disease, but the opposite is true for the dual end points of infection and disease and the triple primary end points when VE_I is low. The proposed multiple-testing method is appreciably more powerful than Bonferroni correction.

In order to investigate the ability of the proposed methods to detect long-term vaccine efficacy, we extended the follow-up time in the above simulation studies from a maximum of 6 months to a maximum of 12 months. We assumed that the event proportions for infection, disease, and severe disease in the placebo group over the 12-month period doubled those of the 6-month period. We reduced all values of vaccine efficacy by 30% to reflect the waning of vaccine efficacy against each end point over time. We tested the null hypothesis that the vaccine efficacy is 0% vs the alternative hypothesis that the vaccine efficacy is greater than 0% at the nominal significance level of 2.5%. The results are summarized in Table 2. Again, the proposed methods can substantially improve statistical power.

DISCUSSION

Here, we present a simple and rigorous framework to consider the totality of evidence when evaluating the benefit of a COVID-19 vaccine in reducing SARS-CoV-2 infection, symptomatic COVID-19, and severe COVID-19. The proposed methods are more robust to different scenarios of vaccine efficacy than the use of a single primary end point. We recommend using the combined test to provide an overall assessment of worthwhile vaccine efficacy, then using the sequential test (Supplementary Appendix 1) to determine the end points against which the vaccine is efficacious.

If a vaccine is more effective in preventing severe than mild COVID-19, then using symptomatic COVID-19 and severe COVID-19 as dual primary end points will be more powerful than using either of the 2 events as a single primary end point. If the vaccine efficacy for infection is nearly as high as that for disease, then using infection, symptomatic COVID-19, and severe COVID-19 as triple primary end points will be the most powerful.

Most phase 3 trials have targeted 90% power for detecting 60% (short-term) vaccine efficacy against COVID-19. The actual power may be lower if the vaccine is less effective, the disease incidence is lower than anticipated, or it is an interim analysis. In our simulation studies, using disease as a single primary end point had only 80% power. However, the proposed methods could boost the power to 90%.

The phase 3 trials under Operation Warp Speed have thus far used symptomatic COVID-19 as the sole primary end point, assessing severe COVID-19 as a secondary end point and assessing a composite burden-of-disease end point as either a secondary end point or an exploratory end point [19]. Under such a plan with a fixed-sequence strategy, hypothesis testing on secondary end points would be permitted only if the result on the primary end point is statistically significant [12]. In the likely scenarios that VE_S is higher than VE_D, using disease and severe disease as dual primary end points will be more powerful than using disease alone as the sole primary end point and thus may accelerate the discovery and deployment of effective vaccines.

Table 2. Statistical Power (%) for Testing the Null Hypothesis of No Vaccine Efficacy Against Infection, Disease, and Severe Disease Over 12 Months

Vaccine Efficacy			Single End Point			Combined Test			Multiple Testing			Bonferroni		
VE_I	VE_D	VE_S	I	D	S	I-D	D-S	I-D-S	I-D	D-S	I-D-S	I-D	D-S	I-D-S
10%	30%	30%	22	84	26	54	80	57	79	78	75	76	77	72
10%	30%	40%	22	84	44	54	85	60	79	80	77	76	79	74
10%	30%	50%	22	84	65	54	89	64	79	84	81	76	83	79
10%	30%	60%	22	84	84	54	92	67	79	90	88	76	89	86
20%	30%	30%	70	84	26	82	80	81	83	78	78	80	77	76
20%	30%	40%	70	84	44	82	85	84	83	80	80	80	79	78
20%	30%	50%	70	84	65	82	89	86	83	84	84	80	83	82
20%	30%	60%	70	84	84	82	92	88	83	90	90	80	89	88
25%	30%	30%	88	84	26	90	80	89	89	78	86	87	77	84
25%	30%	40%	88	84	44	90	85	91	89	80	87	87	79	85
25%	30%	50%	88	84	65	90	89	93	89	84	89	87	83	88
25%	30%	60%	88	84	84	90	92	94	89	90	93	87	89	92
30%	30%	30%	97	84	26	96	80	95	96	78	94	95	77	93
30%	30%	40%	97	84	44	96	85	96	96	80	95	95	79	94
30%	30%	50%	97	84	65	96	89	97	96	84	96	95	83	95
30%	30%	60%	97	84	84	96	92	97	96	90	97	95	89	96

VE_I, VE_D, and VE_S denote, respectively, the vaccine efficacy for infection, disease, and severe disease. I, D, and S denote, respectively, infection, disease, and severe disease. I-D, D-S, and I-D-S denote, respectively, the dual end points of infection and disease, the dual end points of disease and severe disease, and the triple end points of infection, disease, and severe disease. The power pertains to a single test at the 1-sided nominal significance level of 2.5%.

We focused on vaccine trials for populations enriched with high-risk individuals (eg, front-line healthcare personnel, factory workers, older adults, people with underlying health conditions) in whom the risks for infection, disease, and severe disease are all appreciable. In generally healthy populations, such as college students, the majority of infections are asymptomatic and severe disease is rare. For such settings, power can be maximized by using the dual primary end points of infection and disease.

We used Poisson models instead of Cox proportional hazards models for several reasons. First, there are considerable inaccuracies in determining the event times, especially the infection time; the Poisson modeling approach requires only the knowledge of whether or not the event has occurred by the end of follow-up. Second, Poisson models are simpler than Cox models, both conceptually and computationally. Because the event rates are relatively low, the 2 modeling approaches should provide similar results [20]. We fitted both Poisson and Cox models in our simulation studies, and the power of the 2 approaches was nearly identical (Supplementary Appendix 3).

We emphasized hypothesis testing based on score statistics. In Supplementary Appendix 4, we extend our work to general Poisson regression, which can be used to estimate vaccine efficacy, construct confidence intervals, compare multiple vaccines, and accommodate baseline risk factors (eg, age, gender, race, occupation, comorbidity). Baseline risk factors can have a major impact on the occurrences of SARS-CoV-2 infection, symptomatic COVID-19, and severe COVID-19. In addition, some participants in COVID-19 vaccine efficacy trials may become unblinded through the use of available diagnostic tests, and at some point trials may become unblinded. Covariate adjustment in the analysis of vaccine efficacy against end points during post unblinding follow-up is important for minimizing bias due to potential differences in exposure to SARS-CoV-2 between the vaccine and placebo arms.

We developed our methods in order to accelerate the discovery, characterization, and licensure of effective COVID-19 vaccines. An important function of the phase 3 trials is to continue unblinded follow-up of the vaccine and placebo groups after definite evidence of short-term efficacy has emerged. This is done in order to assess the duration of protection and improve precision for assessment of prevention of severe disease as well as for assessment of safety. Duration of vaccine efficacy is an influential parameter in models of population impact of deployed vaccines. An understanding of how vaccine efficacy wanes over time is essential when deciding whether or not booster vaccinations may be required and when estimating the optimal timing of the boosts. The ability of our framework to use the joint distribution of estimators to provide more precise confidence intervals around the 3 vaccine efficacy parameters compared with existing methods (eg, Bonferroni correction) that do not account for the correlation of end points is advantageous regardless of whether 1, 2, or 3 end points are selected as primary.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. The authors are grateful to Yu Gu and Bridget I. Lin for assistance and to 2 referees for constructive comments.

Financial support. This work was supported by the National Institutes of Health grants R01 AI029168, R01 GM124104, P01 CA142538, and UM1 AI068635.

Potential conflicts of interest. D. M. reports employment and stock ownership in Merck & Co, Inc. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Xia S, Duan K, Zhang Y, et al. Effect of an inactivated vaccine against SARS-CoV-2 on safety and immunogenicity outcomes: interim analysis of 2 randomized clinical trials. J Am Med Assoc **2020**; 324:951–60.
2. Jackson LA, Anderson EJ, Rouphael NG, et al. An mRNA vaccine against SARS-CoV-2-preliminary report. New Eng J Med **2020**; 383:1920–31.
3. Mulligan MJ, Lyke KE, Kitchin N, et al. Phase 1/2 study of COVID-19 RNA vaccine BNT162b1 in adults. Nature **2020**; 586:589–93.
4. Folegatti PM, Ewer KJ, Aley PK, et al; Oxford COVID Vaccine Trial Group. Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. Lancet **2020**; 396:467–78.
5. Mercado NB, Zahn R, Wegmann F, et al. Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. Nature **2020**; 586:583–8.
6. Keech C, Albert G, Cho I, et al. Phase 1–2 trial of a SARS-CoV-2 recombinant spike protein nanoparticle vaccine. New Eng J Med **2020**; 383:2320–32.
7. World Health Organization. **2020**. Draft landscape of COVID-19 candidate vaccines. Available at: https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines. Accessed 8 December 2020.
8. Krause P, Fleming TR, Longini I, Henao-Restrepo AM, Peto R; World Health Organization Solidarity Vaccines Trial Expert Group. COVID-19 vaccine trials should seek worthwhile efficacy. Lancet **2020**; 396:741–3.
9. Plotkin SA. Updates on immunologic correlates of vaccine-induced protection. Vaccine **2020**; 38:2250–7.
10. US Food and Drug Administration. **2020**. Development and licensure of vaccines to prevent COVID-19: guidance for industry. Silver Spring, Maryland.
11. World Health Organization. **2020**. WHO target product profiles for COVID-19 vaccines. Available at: https://www.who.int/who-documents-detail/who-target-productprofiles-for-covid-19-vaccines. Accessed 17 September 2020.
12. US Food and Drug Administration. **2017**. Multiple endpoints in clinical trials: guidance for industry. Silver Spring, Maryland. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry.
13. World Health Organization. **2020**. An international randomised trial of candidate vaccines against COVID-19. Available at: https://www.who.int/publications-detail/an-international-randomised-trial-of-candidate-vaccines-against-covid-19. Accessed 20 September 2020.
14. Lin DY, Zeng D. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. Biometrika **2010**; 97:321–32.
15. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics **1979**; 35:549–56.
16. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika **1983**; 70:659–63.
17. Lan KK, Wittes J. The B-value: a tool for monitoring data. Biometrics **1988**; 44:579–85.
18. Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials. CRC Press, **1999**.
19. Mehrotra DV, Janes HE, Fleming TR, et al. Clinical endpoints for evaluating efficacy in COVID-19 vaccine trials. Ann Int Med **2020**; in press.
20. Cuzick J. The efficiency of the proportions test and the logrank test for censored survival data. Biometrics **1982**; 38:1033–9.