STATISTICS IN BIOPHARMACEUTICAL RESEARCH
2022, VOL. 14, NO. 1, 28–32
https://doi.org/10.1080/19466315.2021.1919197

Taylor & Francis
Taylor & Francis Group

Statistical Observations on Vaccine Clinical Development for Pandemic Diseases

Scott Patterson[a], Bo Fu[a], Ya Meng[a], Fabrice Bailleux[b], and Josh Chen[a]

[a]Sanofi Pasteur, Swiftwater, PA; [b]Sanofi Pasteur, Marcy L'Etoile, Lyon, France

ABSTRACT

Traditional vaccine clinical development is an undertaking involving meticulous, multiple studies in multiple populations at risk of infection and disease over multiple years. SARS-CoV-2 and COVID-19 vaccine development is following this traditional development pathway, and accelerated Phase I-II-III clinical programs are being applied. This is not the first time vaccines have been manufactured and tested quickly to meet a public health crisis. Selected statistical concepts pertaining to vaccine efficacy and safety, relevant during the design and implementation of such clinical development programs, will be discussed.

ARTICLE HISTORY
Received November 2020
Accepted April 2021

KEYWORDS
Efficacy; COVID-19; Pandemic; Safety; SARS-CoV-2; Vaccine

1. Introduction

Pandemic disease is a recorded reality of human history (Armstrong 2016; McNeill 1976). In the past, generational development of immune system memory and resistance (sometimes called "herd" immunity) was the path to disease eradication. Modern medicine, however, has resulted in treatments and in vaccines which have ameliorated (e.g., polio, chicken-pox, and influenza) or eliminated (e.g., small-pox) disease vectors and the resulting disease.

Ten years and approximately a billion dollars is the usual time and expenditure to develop a safe and effective and well manufactured product in the modern, Western regulatory environment (DiMasi 2001; DiMasi, Hansen, and Grabowski 2003). Most of this time is spent performing clinical trials. Even if scientifically successful (subject to local regulatory and governmental requirements), only certain selected products may subsequently be recommended for use in large segments of local human populations.

Recent notable infectious disease pandemics involving development of vaccines as a measure to support public health include H1N1 influenza (European Centre for Disease Prevention and Control 2012), Ebola (Branswell 2020), and most recently SARS-CoV-2 (the virus which causes COVID-19 disease) (World Health Organization 2020), but these are supplemented with other notable failures (e.g., HIV).

Our focus in this work is on selected statistical observations concerning vaccine efficacy and clinical safety testing for vaccines developed during pandemics or epidemics, not exclusively on the SARS-CoV-2 pandemic. This article summarizes some of our observations on these statistical matters as these aspects of vaccine clinical development are not very well known amongst trial statisticians. Readers interested in a more comprehensive review of vaccine efficacy for SARS-CoV-2 may find (Hodgson et al. 2021) of interest.

In the next section, selected observations on the statistical aspects of vaccine efficacy testing will be reviewed followed by a section on clinical safety testing. In the discussion section, we include some comments on the status of SARS-CoV-2 vaccine development as of the time of writing of this work; however, the field is moving very rapidly, and discussion of current events is necessarily limited in scope.

2. Observations on Vaccine Efficacy Testing

2.1. Statistical Methods

In brief, let x_1 be the number of cases on control $i = 1$, and x_2 be the number of cases on vaccine $i = 2$. Then, we know that $x_i \sim \text{Bin}(n_i, p_i)$ where n_i is the number of subjects exposed in each treatment group and p_i is the probability of interest and *Bin* denotes the Binomial distribution so that $x_i \sim P(\lambda_i)$, where P is the Poisson distribution and $\lambda_i = n_i(p_i)$ when n_i is large and p_i is small. Let $R = \frac{\lambda_2}{\lambda_1}$, such that Vaccine Efficacy VE $= 1 - R$. Readers may also find Nauta (2010) useful for understanding the estimation of vaccine efficacy.

Case size derivations summarized in Table 1 are described in Chow, Shao, and Wang (2003) and are derived assuming the binomial distribution of the number of cases in vaccine group is conditioned on the target total number of cases. The Bayesian derivations given in Table 2 are based upon Chu and Halloran (2004) and are applied on the proportion of cases in vaccine group conditional on the total number of cases (here denoted p). The density distribution for the posterior distribution for p is derived from combination of a prior (beta(1,1)) and the likelihood of the observed data as given in the table by numerical integration of the density distribution of the parameter p in SAS.

2.2. Impact of Imperfect Tests and Mutation of the Virus

The identification of a case depends upon testing of biologic samples with assays. Cases may begin to present based on

CONTACT Scott Patterson scott.patterson2@sanofi.com Sanofi Pasteur Inc, Biostatistics, 1 Discovery Drive, Swiftwater, PA 18370.

Table 1. Required minimum number of cases for each combination of target VE (VEt) and lower confidence bound acceptance limit (VE0) with Type I error rate $\alpha = 2.5\%$ and Power=90%.

	VEt		
VE0	50%	60%	70%
0%	99	61	37
10%	139	74	43
20%	216	105	59
30%	419	160	78

Table 2. Posterior probability that true VE $\leq$ 0 given that vaccine efficacy is observed in the study with lower confidence limit > VE0.

Design VE0	Number Cases	Case Split Vax:Pbo	Observed LB	Probability VE $\leq$ 0
10%	74	26:48	10.9%	0.005
20%	105	36:69	20.8%	0.0006
30%	160	53:107	30.5%	0.000009

clinical diagnosis—for example, cough, fever. Tests are then run (e.g., X-rays, biological samples) to verify disease state, and samples are then assayed to determine the cause of the disease state. A certain fraction of the clinical diagnoses denoted $0 \leq m \leq 1$ will turn out to be driven by the pathogen of interest (i.e., the one targeted by the vaccine). Of these, the assays will identify whether they are vaccine-type (VT), or not. This factor m is generally derived from epidemiology data.

Specificity and sensitivity of the assays should be accounted for in this setting. Assay specificity is the assay's rate of identification of true negatives—that is if the sample is truly negative for the VT pathogen, the assay correctly determines it as such. We will refer to this as $0 \leq t_n \leq 1$. Assay sensitivity is the assay's rate of identification of true positives—that is, if the sample is truly positive for the VT pathogen, the assay correctly determines it as such. We will refer to this as $0 \leq t_p \leq 1$.

It is expected that both t_p and t_n lie close to 1 for use in a vaccine clinical efficacy (VE) study. It is desirable that they indeed are 1, because if not, then the observed VE in a clinical trial will be blunted by these factors such that

$$\mathrm{VE}_{\mathrm{OBS}} = \frac{\mathrm{VE}}{1 + C}$$

where

$$C = \frac{(1 - m)(1 - t_n)}{m(t_p)}$$

where VE is the true vaccine efficacy, and $\mathrm{VE}_{\mathrm{OBS}}$ is the observed vaccine efficacy in the clinical trial blunted for less than perfect specificity (Lachenbruch 1998). For example, if $m = 0.1$ with $\mathrm{VE} = 0.9$, $t_p = 0.99$, and $t_n = 0.995$, then observed vaccine efficacy will be expected to be $\mathrm{VE}_{\mathrm{OBS}} = 0.86$.

The factor m however cannot be regarded as a constant except in the context of a single clinical trial of limited duration. Indeed, as multiple vaccines and treatments become available and requirements for public health control (e.g., masks, social distancing) are decreased, the role of competing respiratory illness would be expected to increase leading to decreased m. Similarly, stability of the factors m, t_n, and t_p also depends upon the key assumption that the virus does not mutate significantly. In such a case, the fraction of VT cases would be expected to decrease, even if the vaccines being developed remain at least partially effective against mutated strains, thereby decreasing m and further blunting the estimates of vaccine efficacy. Additionally, the t_n and t_p for the mutated virus and the disease it causes would also be expected to be lower until assays are upgraded to detect the mutation.

Consider a situation where VE is smaller (0.50) against a mutant strain. The additional complicating factors of decreased $m = 0.05$ with $t_p = 0.99$ and $t_n = 0.97$ for the competing stain and assay would result in blunting of the observed VE in the study to $\mathrm{VE}_{\mathrm{OBS}} = 0.32$. In practice, m, t_n, and t_p would be determined before a study begins, with the effect on VE adjusted for by increasing target case count and-or sample size; however, rapid mutation of the virus and changes in medical practice during the study can clearly affect the resulting estimates.

2.3. Lower Confidence Limit Acceptance Bound

Article 351 of the USA's Public Health Service Act defines the criteria necessary to license a vaccine (Gruber 2014).

- Data must show the vaccine candidate is "safe, pure, and potent."
- "Potent" has been interpreted to include vaccine efficacy (prevents or lowers disease incidence).

No statute or regulation requires a specific minimum level of vaccine efficacy, but in vaccine research and development programs, these legislated factors are combined with the addition of mandatory control of the Type 1 error rate and the ethics of prevention in vulnerable populations to establish acceptance criteria. USA FDA, for example, requires only one vaccine efficacy study for registration, but the lower bound of the vaccine efficacy confidence interval (denoted LB) must be "acceptably" better than 0. "More than one study may be necessary to substantiate findings, especially if LB is close to 0 (greater likelihood of a type 1 error)" (Gruber 2014).

Another key factor in the choice of LB is driven by the power to demonstrate VE and case numbers required. A higher LB increases the number of cases required (see Table 1).

Acceptance boundaries typically vary from roughly 10% to 20% based on our experience for vaccine efficacy, depending upon the true degree of efficacy and a number of other factors, but they may be as high as 50% in some very unusual circumstances (USA Food and Drug Administration 2008). As shown in Table 2 (last column), even with a 10% LB, the probability of false positive VE result is still low with an acceptable range (0.005 for true VE $\leq$ 0); for a 20%–30% LB, this probability is much lower.

3. Observations on Vaccine Safety

3.1. Statistical Methods

In keeping with O'Neill (1988), untoward events following vaccination are regarded as binomial (Bin) distributed such that:

$$e_i \sim \mathrm{Bin}(n_i, p_i)$$

where e_i is an event of interest in treatment group $i = v, c$, where v denotes vaccine and c denotes the control group, n_i denotes the

sample size, and p_i is the unknown probability of an event. Here, it is assumed that untoward events e_i have $p_v > p_c$, and that such events are not anticipated following vaccination. Some other events, local reactions and systemic events, are expected when vaccine prompts a response from the immune system, and those are not in the scope of this work. Notation has been changed from x_i in the previous section to denote that unlike efficacy testing, safety testing is not powered in clinical trial design.

Alteration of the immune system can be regarded as permanent for the purposes of this assessment, meaning that once vaccinated, the immune system and related body processes do not return to basal status. In general, regulations require at least a 6-month safety period of assessment following the last vaccination in at least 3000 subjects to build a safety database sufficient to be considered for vaccine approval to market. Most rates of untoward experiences are quoted in figures such as 1 event per some multiplier of thousand subjects per year. Hence, without loss of generality, the probability of an event should be regarded as 1/2 the background rate for subjects followed for 6 months in the control group.

Of interest is the signal detection and statistical interpretation of untoward, unanticipated events when $p_v > p_c$ for small p_i considering the observed n_i. Statistical tests for this are numerous, and for the purposes of this work the Fisher's Exact test and Newcombe's hybrid score were used. In this setting n_i is predetermined to provide adequate power to assess the study's primary efficacy objective, not to test $p_v = p_c$; hence, while the studies are randomized and data hence informative for causal inference, the detection of false positives in safety data is to be expected. Type I error control in this setting has been discussed by others (e.g., Nauta 2010), but as our interest is in quantifying uncertainty around the observation and testing of increasingly rare untoward events in Phase 3 studies, we will not explore Type I error control further in this work.

One thousand studies of sample size n_i=10,000–40,000 were simulated using SAS for rare to increasingly rare untoward events to study the probability of observing such an event, and if so, whether standard statistical tests can differentiate an effect from noise. Simulation findings are summarized in Table 3.

3.2. Likelihood of Observation of Untoward Events in Clinical Trials

Exacerbated respiratory disease (ERD), noted of concern (FDA Guidance 2020a), does not appear to be such a rare event. Aspirin ERD occurs at a rate of 0.3%–0.9% each year in the general population with higher rates in patients with asthma (Li, Lee, and Abuzeid 2019). Assuming the vaccine causes such an event with similar probability and subjects are followed for 6 months, we should observe at least one such occurrence in in a sample size of 10,000–40,000 subjects on vaccine, and the likelihood of having statistically interpretable result using the Newcombe Hybrid Score confidence interval or Fisher's exact test with an equally sized placebo group is high. This assumes the probability of an event is very small in control subjects (0.01% for these simulation studies).

However, consider the recent experience with intussusception in infants when given vaccines for Rotavirus. The first such vaccine was withdrawn from the market when it was estimated that vaccination resulted in an excess of 1 case per 5000 subjects vaccinated on a background rate of 1 naturally occurring case per 2000 infants per year (Centers for Disease Control and Prevention 2020a). According to the CDC, the vaccines being used now also increase intussusception, but at a lower rate (1 additional case in 20,000–100,000 vaccinated subjects per year) (Centers for Disease Control and Prevention 2020b). Using 1 case in 4000 subjects $p_c = 1/4000$ as our background (control) probability for a theoretical untoward event in a group of subjects followed for 6 months, at least one such event will be observed using simulation as greater than 90% of studies. Correspondingly, assuming vaccination increases the probability of an event such that $p_v = (1/4000) + (1/10,000)$ (again assuming subjects are followed for 6 months), the estimate how often such at least one such event will be observed using simulation as 97%–100%. However, the probability of statistically differentiating this increase is unlikely, even with a very large sample size of 40,000 subjects per group.

Consider a disease like Guillain-Barre syndrome (GBS) where the background rate is far lower (six cases per million subjects per year (NIH US National Library of Medicine 2020)), then the probability such at least one such event will be observed in a sample size of 10,000–40,000 subjects is only 3%–12%. Even if this probability of an event doubles following vaccination, the likelihood of observing such an event in 10,000–40,000 subjects remains low (6–22%), but the likelihood of statistical testing identifying the increase is not to be expected.

Here, we were concerned with the probability of statistically detecting that $p_v > p_c$. As shown in Table 3 this probability is low, and as events become increasingly very rare statistically significant findings should not be expected, if indeed such

Table 3. Summary of simulated untoward events and probability of statistical differentiation.

N/Group	p_c	$P(e_c \geq 1)$	p_v	$P(e_v \geq 1)$	NHS	Fisher
10,000	0.0001	62.0%	0.0025	100%	99.9%	99.9%
20,000	0.0001	85.7%	0.0025	100%	100%	100%
30,000	0.0001	95.0%	0.0025	100%	100%	100%
10,000	0.00025	90.3%	0.00035	97.2%	4.2%	2.1%
20,000	0.00025	99.5%	0.00035	99.9%	7.1%	4.4%
30,000	0.00025	100%	0.00035	100%	8.1%	6.0%
40,000	0.00025	100%	0.00035	100%	11.1%	7.7%
10,000	0.000003	2.8%	0.000006	5.8%	0%	0%
20,000	0.000003	6.5%	0.000006	10.8%	0%	0%
30,000	0.000003	8.5%	0.000006	17.0%	0%	0%
40,000	0.000003	12.0%	0.000006	21.9%	0%	0%

NOTES: 1000 simulation studies per scenario. p_i = Prob. of an event on $i = v$ vaccine or $i = c$ control. $P(e_i \geq 1)$ = Percent of studies one or more events observed. NHS: Percent of studies with Newcombe Hybrid Score 95% CI > 0. Fisher: Percent of studies with $\hat{p_v} > \hat{p_c}$ and Fisher's p-value < 0.05.

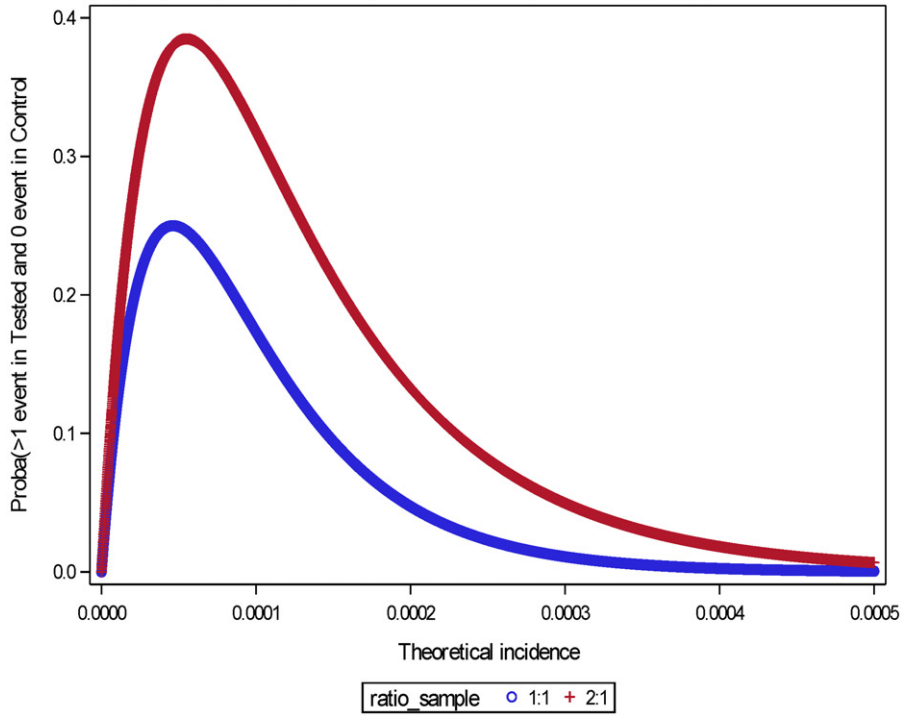

Figure 1. Probability to observe at least one event in Vaccine group and 0 event in Control group as a function of true and equal incidence of the event for 1:1 and 2:1 vaccine:control randomization ratios

increasingly rare untoward events are observed in clinical trials at all.

The previous work (Patterson et al. 2020) has studied the impact of imbalanced randomization on the statistical tests. While the statistical tests applied here remain unbiased in imbalanced designs, findings in this work indicate that statistical testing is unlikely to be informative for rare untoward events, and interpretation of estimates of p_v relative to p_c is complex as rare events are more likely to be observed in the larger trial arm—see, for example, Figure 1.

4. Discussion

The global COVID-19 pandemic caused by a novel coronavirus SARS-CoV-2 may be ameliorated or stopped by safe and effective vaccines. Unprecedented collaborations have arisen among academia, industry, and governments around the globe, and much progress have been made to accelerate vaccine research and development. Multiple vaccine candidates from various platforms are currently in clinical development, and several vaccines have been authorized for emergency use. In this article, we discuss three key statistical issues that are important in establishing safety and effectiveness of a COVID-19 vaccine, based on current understanding of the science and available regulatory guidance.

Given the complicating factors discussed in Section 2, strict interpretation of vaccine efficacy data relative to a VE target would appear to be unwise in the context of a mutating virus and imperfect assays as looks to be the case for SARS-CoV-2. As a practical matter it will be difficult to compare VE estimates that differ in time of study or duration of study as the virus mutates, particularly if the estimates of VE involve differing endpoints and assays. Given the number of vaccine candidates in development, the use of a 30% lower confidence bound acceptance limit also appears conservative and will not allow for latitude in differentiation in importance amongst mutating strains.

Given the very short duration of follow-up in clinical trials for Emergency Use, it is unlikely that rare untoward events caused by the vaccine will be observed in clinical studies, and it is even more unlikely that such events will generate sufficient data to be statistically interpretable. In all likelihood then, the observation of such events, quantification of risk, and benefit:risk assessment will follow emergency use of vaccines in a broader population. This highlights the importance of unbiased, accurate real-world surveillance, data, and evidence generation. Some systems are in place to monitor vaccine safety in the real world (e.g., Zhao et al. 2020). However, as evidenced by recent announcements about the pending development of real-world evidence guidance in the United States, China, Europe, and the United Kingdom, full understanding of how to interpret such real-world data for regulatory decision-making is not yet available. It is important to ensure unbiased and accurate estimates are obtained from such systems, but that falls beyond the scope of this article (and is likely to be a very hot research topic moving forward).

It is important to note that there are a number of other factors which have not been studied (thus far) in these accelerated COVID-19 Phase I-II-III vaccination development clinical trials including: duration of protection; the impact of co-administration of other vaccines (e.g., influenza) on COVID-19 vaccines (and vice versa); impact of co-administration with acetaminophen and ibuprofen on performance of the COVID-19 vaccines; use of the COVID-19 vaccine in special populations

(e.g., pediatrics, asthmatics); when (if) to give a booster dose; lot consistency. These are the type of additional clinical studies that take a large part of the 10 years usually required to develop and register a vaccine and, presumably, are being saved for postemergency commitments.

It is also to be hoped that sufficient VE studies are being performed using accurate, precise, and calibrated immunogenicity assays. As data become available, this may allow for identification of a threshold or correlate of immunogenic protection. If such a level can be identified or a correlate can be established, immuno-bridging studies would suffice to address efficacy for subsequent vaccine candidates which may use different platforms other than that studied in the previous efficacy trials. Eventually, the established immune biomarkers regardless of platforms would predict protection. This alternative approach could be an important option if conducting new placebo-controlled efficacy trials following emergency use of highly efficacious COVID-19 vaccines is difficult. Other adaptations could include leveraging real-world data to support expanded use. Eventually, it is not one size fits all; a global pandemic requires creative approaches that adapt to growing knowledge of the novel virus and disease, mutations, treatments, and vaccines.

Researchers, clinicians, and regulators are still on learning curve to understand the science of this novel SARS-CoV-2 virus and COVID-19 disease. Many uncertainties remain, especially around how the virus evolves, the epidemiology, human immunology against the virus, and clinical progression in patients. Improving knowledge could guide appropriate approaches to develop safe and effective COVID-19 vaccines. As statisticians should know very well, "absence of evidence does not equate to evidence of absence" (Altman and Bland 1995), and we currently have a great deal of the "absence of evidence" for SARS-CoV-2. There will be many factors to study in clinic for COVID-19. The clinical testing of COVID-19 vaccines will not end in 2021 and is likely to be very complex and contentious. And, COVID-19 is far from the first, and will not be the last, infectious disease world health crisis.

Acknowledgments

At the time of the preparation of this work, all authors were employees of Sanofi Pasteur. SP, BF, YM, FB, and JC are shareholders of Sanofi.

This article represents the views of the authors and does not necessarily represent the views of current (and former) employers, their divisions, their personnel, nor any affiliates.

References

Altman, D.G., and Bland, J.M. (1995), "Absence of Evidence Is Not Evidence of Absence," *British Medical Journal*, 311, 485. [32]

Armstrong, D. (2016), *The Black Death: The World's Most Devestating Plague*, The Great Courses. [28]

Branswell, H. (2020), "'Against All Odds': The Inside Story of How Scientists Across Three Continents Produced an Ebola Vaccine," available at *https://www.statnews.com/2020/01/07/inside-story-scientists-produced-world-first-ebola-vaccine/* [28]

Chow, S. C., Shao, J., and Wang, H. (2003), *Sample Size Calculations in Clinical Research*, New York: Marcel Dekker. [28]

Chu, H., and Halloran, M. E. (2004), "Bayesian Estimation of Vaccine Efficacy," *Clinical Trials*, 1, 306–314. [28]

Centers for Disease Control and Prevention (2020a), "Rotavirus Vaccine (RotaShield®) and Intussusception," available at *https://www.cdc.gov/vaccines/vpd-vac/rotavirus/vac-rotashield-historical.htm* [30]

Centers for Disease Control and Prevention (2020b), "Questions & Answers About Intussusception and Rotavirus Vaccine," available at *https://www.cdc.gov/vaccines/vpd/rotavirus/about-intussusception.html* [30]

DiMasi, J. (2001), "New Drug Development in the USA From 1963 to 1999; Risks in New Drug Development – Approval Success Rates for Investigational Drugs," *Clinical Pharmacology and Therapeutics*, 69, 286–307. [28]

DiMasi, J., Hansen, R., and Grabowski, H. (2003), "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics*, 22, 151–185. [28]

European Centre for Disease Prevention and Control (2012), "Narcolepsy in Association With Pandemic Influenza Vaccination," available at *https://www.ecdc.europa.eu/sites/portal/files/media/en/publications/Publications* [28]

FDA Guidance (2020a), "Development and Licensure of Vaccines to Prevent COVID-19," available at *http://www.fda.gov/cder/guidance/* [30]

Gruber, M. (2014), "Regulatory Considerations for Determining Vaccine Efficacy: USA FDA Perspective," *Global Vaccine and Immunization Research Forum*. [29]

Hodgson, S., Mansatta, K., Mallett, G., Harris, V., Emary, K., Pollard, A. (2021), "What Defines an Efficacious COVID-19 Vaccine? A Review of the Challenges Assessing the Clinical Efficacy of Vaccines Against SARS-CoV-2," *Lancet Infectious Disease*, 21, e26–35. [28]

Lachenbruch, P. (1998), "Sensitivity, Specificity, and Vaccine Efficacy," *Controlled Clinical Trials*, 19, 569–574. [29]

Li, K., Lee, A., and Abuzeid, W. (2019), "Aspirin Exacerbated Respiratory Disease: Epidemiology, Pathophysiology, and Management," *Medical Sciences*, 7, 45. [30]

McNeill, W. H. (1976), *Plagues and Peoples,* New York: Anchor Books. [28]

Nauta, J. (2010), *Statistics in Clinical Vaccine Trials*, Springer: London. [28,30]

NIH US National Library of Medicine (2020), "Guillain-Barre Syndrome," available at *https://medlineplus.gov/genetics/condition/guillain-barre-syndrome/* [30]

O'Neill, R. (1988), "Ch 13. Assessment of Safety," in *Biopharmaceutical Statistics for Drug Development*, ed. K. Peace, 543–604. New York: Marcel Dekker. [29]

Patterson, S., Bailleux, F., Chen, J., and Zhu, M. (2020), "Imbalanced Randomization in Vaccine Clinical Safety Trials," *Statistics in Biopharmaceutical Research*, 12, 279–283. [31]

USA Food and Drug Administration (2008), "Statistical Review and Evaluation – Rotarix," available at *https://www.fda.gov/vaccines-blood-biologics/vaccines/statistical-review-and-evaluation-rotarix* [29]

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* [28]

Zhao, L., Lee, S., Li, R., Ong, E., He, Y., and Freed, G. (2020), "Improvement in the Analysis of Vaccine Adverse Event Reporting System Database," *Statistics in Biopharmaceutical Research*, 12, 303–310. [31]