Some Statistical Methods to Accelerate Covid-19 Vaccine Testing

Jorge Luis Romeu, Ph.D.
https://www.researchgate.net/profile/Jorge_Romeu
http://web.cortland.edu/romeu/; Email: romeu@cortland.edu

1.0 Introduction

Much has been written and said, in the last few weeks, about accelerating the phases of Covid-19 (i.e. development and testing) vaccine clinical trials[1] to produce a working product. For example:

> Trump Administration's Operation Warp Speed Accelerates AstraZeneca COVID-19 Vaccine[2]**:** *The agreement between AstraZeneca and the Biomedical Advanced Research and Development Authority (BARDA), part of the HHS office of the Assistant Secretary for Preparedness and Response,* ***will accelerate the development and manufacturing of the company's investigational vaccine to begin Phase 3 clinical studies*** *this summer with approximately 30,000 volunteers in the United States.*

Clinical trials length is closely related to the number **n** of participating volunteers. Determination of said *sample size "n"* required in testing and confidence interval (CI) derivation has always been of great importance. For, sampling is both expensive and time consuming, and sample size *carries a big price tag in time, resources or both*. In clinical trials *time is crucial,* and the number **n** of *volunteers* is *scarce*, and the *pressure* to find/release a vaccine, is *strong*.

When *samples are taken all at one time*, it is called single (or fixed) sampling. An alternative consists in taking the *samples in multiple stages*, and assessing their results at every stage. This allows the possibility of stopping the process and reaching an early decision, if certain conditions are met. For, if the data show a clear-cut trend in favor of (or against) the hypothesis being tested, from the start of the clinical trial, then shortening the test can save significant time and resources. In such cases the *samples are taken in successive stages,* according to the assessment of results obtained from the previous sampling stages. This is known as "multiple sampling".

Thence, there are situations where it is more efficient to *take samples sequentially*, as opposed to all at one time, and to *define a stopping rule* to terminate the sampling process. Taking samples sequentially and assessing their results at each stage allows for the possibility of stopping the process and *reaching an early decision*. If the situation is clear-cut favorable or unfavorable (for example, if the sample shows that a vaccine is definitely effective or poor), then terminating the process early *can save significant time and resources*. In the case *where results are ambiguous,* and we require additional information to take a better decision, *we continue sampling*.

1. _____

[1] https://www.cancer.org/treatment/treatments-and-side-effects/clinical-trials/what-you-need-to-know/phases-of-clinical-trials.html describes clinical trials, their phases and objectives/components

[2] https://www.hhs.gov/about/news/2020/05/21/trump-administration-accelerates-astrazeneca-covid-19-vaccine-to-be-available-beginning-in-october.html discusses government procedures to accelerate the clinical trial process.

Clinical trials are about assessing the *success or failure* of a treatment, medication or vaccine. Thence, we present several *industrial sampling plans* that discuss *test for attributes* (pass/fail data) that *follow the Binomial distribution*). We then explore *double and multi-stage sampling* plans. We continue with a discussion of *Expected Sample Number* (ASN) a performance measure used *to assess efficiency* of multi-stage sampling plans. We conclude with sequential probability ratio tests (*SPRTs*). We illustrate our discussions via numerical and practical examples.

In References 1 and 2, double sampling plans are discussed and extended to higher dimension plans, namely sequential tests. In References 3, the problem of calculating the sample size **n** for experimentation is discussed. In References 4 and 5, samples for acceptance testing and censored data are discussed. A useful industrial statistics textbook is given in Reference 6.

2.0 Clinical trials for vaccine development and statistical hypothesis testing

A clinical trial entails, implicitly, the testing of a statistical hypothesis. We want to establish with high confidence whether, say a drug or a vaccine, has (or doesn't have) an effect on an illness or disease. Therefore, we need to first discuss hypothesis testing under such context.

In hypothesis testing, we define the Null (H_0) hypothesis that expresses the status quo (in this case that the drug/vaccine *has an effect*). Then, we define the Alternative (H_1) hypothesis as the negation of the Null (that the drug/vaccine actually has *no effect*). We then take a sample of pre-established size "n" of the *random variable X* (in this case, the result of a patient taking said drug or vaccine). And based upon the results from such sample (say, of fixed size n), we take a decision regarding these two hypotheses. This is the single stage sampling procedure.

Random *variable X is Bernoulli*, for it has *two outcomes*: *success* (X=1) with probability p, or *failure*, (X=0) with probability (1-p) and $0 \leq p \leq 1$. Define X=0 if the drug/vaccine has an effect (in the vaccine case, it protects the patient); and X=1, if the drug or vaccine does not works (it does not protect the patient, who gets infected with Covid-19). *The sum* of the "n" Xs (*number of patients* who get infected with Covid-19) is Binomial (n, p). And based upon this sum we take the decision of rejecting or not the Null hypothesis (H_0) that the drug or vaccine has (or has not) had an effect, with a given probability of committing an error taking such decision.

We can place a value on the *acceptable effect* of said drug or vaccine. For example, we say it is *acceptable* if the vaccine fails to work in less than 10% of cases ($p \leq 0.1$); and it is *unacceptable* if it fails in more than 20% of cases ($p \geq 0.2$). *Alpha* (α) is the *probability of* committing the *Type I Error*: *rejecting H_0 when it is true*; and *Beta* (β) is the *probability* of *accepting H_0 when it is false*. In medical terms, *Alpha* (α) is the *probability* of stating that *the drug or vaccine does not have an effect* on the disease, *when it actually has one*, and thence, withdrawing the release of a safe drug/vaccine to the public when it works well. *Beta* (β) *is the converse* probability.

We want low probabilities of errors, say 5% for Alfa ($\alpha=0.05$), and 10% for Beta ($\beta=0.1$). The required sample size "n" for this hypothesis test depends, in addition to α and β, on the natural variability (σ), and on how far apart the values p for H_0 and H_1, are. Then, "n" is obtained as:

$$n = \left[\frac{(z_\alpha + z_\beta)\sigma}{\delta} \right]^2$$

Where Zs are the standard normal table values for α and β, and $\delta = p_1 - p_0$. The sample size "n" for a test that detects difference $\delta = 0.2-0.1 = 0.1$ *with errors* $\alpha = 0.05$ *and* $\beta = 0.1$, *when* $\sigma = 0.6$ *is:*

$$n = \left[\frac{(z_{0.05} + z_{0.1})\sigma}{\delta} \right]^2 = \left[\frac{(1.65 + 1.28) \times 0.6}{0.1} \right]^2 = 309$$

This means that we would need 309 patients for the clinical trial. But this is the beginning of our most serious problem: *how to detect that an effect has occurred, and is due to the drug/vaccine*.

In industrial statistics, this is straight-forward. Say, you receive a batch of light bulbs, take 309 of them at random, and test them. If a light bulb works, it will light up; if it does not it will not. But in a clinical trial, things work differently, and we need to adapt these procedures accordingly.

To start, we need to *define when the drug/vaccine works.* Say it is a drug. *Works* means that the pain is relieved in 48 hours; the fever is down in 24 hours, cough is gone in three days, etc. Then, some people will obtain relief naturally, because of their immune system, pure luck, etc., that has nothing to do with the drug being tested. Other patients will be very sick, or have a poor immune system, etc. and will not get better even with a magical medicine! It is difficult to establish if the effect observed on the patient is coming from the drug, or is coming from another source.

A *sample* (volunteers, available patients, etc) is obtained and *divided at random into two groups*. *One group* will be given *the drug/vaccine*, and *the other, a "placebo"* (nothing). Patients will not know what is been given to them. Doctors will not know whether they are administering the real medication or the placebo (vaccine or placebo) to the patients. *Such tests are known as Doubly Blinded clinical trials*. However, there is always a record of such assignment in the clinical trial protocol that can be accessed by a second, independent group of researchers, and can be used to conduct additional statistical procedures, as suggested in our next section.

If, say, the vaccine works, the percent "p_1" of *inoculated* patients that will become infected with Covid-19 will be much smaller (say k times) than the percent "p_2" of those in the *Placebo group*. This may be *assessed using a two-sample t-test* (with H_o: $p_1 \leq kp_2$ v. H_1: $p_1 > kp_2$). Establishing such difference *for a drug* is easier, because a *disease* can be *diagnosed beforehand*, and the *sample* can be drawn from *patients that suffer* from said *disease*. *With a vaccine things are very different*, as we deal with *virus infections* that are *not yet present*. Let's see how it works:

Non-infected individuals are recruited into a clinical trial for a vaccine, and then divided into two groups, at random. Some will get the vaccine and others, the placebo, in a doubly blinded clinical trial. Then *they need to get infected*, to test the effectiveness of the vaccine. Thence, they need to be *exposed to the virus* either *by inoculation, or by interaction* within an infected environment.

If the sample is small, and vaccinated volunteers are willing to be inoculated with the virus, this will be done (in earlier phases). Otherwise, we need to release the vaccinated volunteers within an infectious environment and wait to see (or hope) that they become infected with the virus.

The first method is the most efficient, because *we know* that *patients have been inoculated with the virus*. We can then *define a period of time T* such that, if the vaccinated and infected patient does not develop Covid-19, *it can be assumed that the vaccine has done its job*.

When a large number of patients are released in an infectious environment, some will become infected, and some will not. The number of infections in patients receiving the placebo should also be larger, and the hypotheses may also be tested using a two-sample t-test. But the number of infections will depend on the number of patients and length of the clinical trial. If we curtail these, the number infected may be too small to establish a difference with a good probability[3].

3.0 Multiple (double, triple etc.) stage hypothesis testing

If results are clear, in favor or against a drug/vaccine, then there is no problem. But if results are too close to the limits established say, if testing for H_0: $p \leq 0.05$ we obtained p=0.048 or p=0.052, then we need to test further, with a larger sample, or take a longer test time. In statistics this can be improved by implementing multiple (double, triple etc.) stage testing procedures.

The logic behind double (multiple) sampling schemes is that, if initial results are clear-cut good or bad, we take a decision based on the first sample only. If there are some doubts, then we draw a second (additional) sample, and collect more information to be used in reaching a decision. This other method lowers the risk of taking the wrong decision, at the cost of a longer and more expensive process (i.e. drawing the second sample will cost more time and money).

Let's illustrate via an example that shows how to build such a test procedure. Assume we have a vaccine that requires (H_1) protecting at least 90% of immunized, with confidence $1-\alpha=0.95$. To test this vaccine, we inoculate "n" say 20 immunized with this virus, wait for time T, and then count how many of them are infected. The *number of infections* **X** is again *distributed Binomial with* $\boldsymbol{n=20}$, and where $\boldsymbol{p}$ is the *probability of infection of the immunized* using the vaccine, and **c** is the *number of infected* patients in the clinical trial. We express such Binomial probability as:

$$P\{X > c\} = 1 - P\{X \leq c\} = 1 - \sum_{x=1,c} C_x^n \, p^x (1-p)^{n-x}$$

If the vaccine *protects 90% or more* of those receiving it (H_0: $p \leq 0.1$), the *vaccine* is considered *efficient*, in which case it will be used. If the vaccine protects 80% or less of patients receiving it (H_1: $p \geq 0.2$) the *vaccine* will be *discarded*. We then define a *double sampling plan* that analyzes a *first sample of size* $n_1=20$ of inoculated patients, and then *counts the number of virus infections* **X** *in the sample, after time T:* S ($n_1=20$, $n_2=20$, $c_1=14$, $c_2=15$, $c_3=33$). If $X > 15$, we reject H_0 (and assume H_1). If $X < 14$ we don't reject H_0 (the drug is helpful). *If* $14 \leq X \leq 15$, *a decision would be too uncertain*. So we *draw a second sample of size* $n_2 = 20$, *and count the number of survivals (Y)* in this second sample. Then, *if* $X + Y < 33$ *we don't reject* H_0 (and decide that the new *vaccine* is acceptable). But if $X + Y \geq 33$ *we reject* H_0; assume H_1 and decide that the new *vaccine* is not efficient and should not be released.

1.

[3]There is a procedure in industrial statistics to estimate the sample size "n" required, in order to obtain, with a given probability, a minimum number of failed cases. See Reference 3, in the Bibliography, for more details.

The *probability of acceptance* for such double sampling plan S, for any p, is given by:

$$P\{Accepting.Drug\} = P\{Accepting.Initially\} + P\{Initially.Inconclusive.Then.Accept.at.2nd\}$$
$$= P\{First.Successes \geq 16\} + P\{First.Successes = 14.or.15.and.Combined.Successes \geq 33\}$$
$$= \sum_{x=16}^{20} Bin(x; n = 20, p) + Bin(x = 14; n = 20, p) \times [Bin(x = 19; n = 20, p) + Bin(x = 20; n = 20, p)]$$
$$+ Bin(x = 15; n = 20, p) \times [Bin(x = 18; n = 20, p) + Bin(x = 19; n = 20, p) + Bin(x = 20; n = 20, p)]$$
$$= 0.957 + 0.0089 \times (0.2701 + 0.1215) + 0.0319 \times (0.2851 + 0.2701 + 0.1215) = 0.982$$

In general, to find the required *S (n, c) plan parameters* we establish a system of two Binomial equations that fulfill such required Types I and II errors (or risks) of the hypothesis test problem:

$$\sum_{x=0}^{c} C_x^n p_0^x (1-p_0)^{n-x} = 1-\alpha; and: \sum_{x=0}^{c} C_x^n p_1^x (1-p_1)^{n-x} = \beta^*$$

Solving this system of two equations, we obtain the appropriate values of "c" and "n" for the required S(n,c) Plan. For several stages, we repeat the procedure, shown above. Alternatively, the ANSI/ASQC Z1.4 (1993)[4] quality manual sampling tables can be consulted.

When the sample size is large (n > 20) the random variable **Number of Failures** approximates the Normal, with $\mu = np$ and $\sigma^2 = np(1\text{-}p)$. We can then, using the same two hypothesized p_i, for i = 0,1, and the two errors or risks α and β given above, establish the system of two simultaneous equations below, and find adequate values for both n and c of the required S(n,c) Plan:

$$\frac{c - np_0}{\sqrt{np_0(1-p_0)}} = z_\alpha;$$
$$\frac{c - np_1}{\sqrt{np_1(1-p_1)}} = -z_{\beta^*}$$

Here, the z_α are the Normal Standard percentiles for probability α. Solving this system for "n" and "c", we obtain the equations that will yield the sample size 'n' and the critical number 'c' fulfilling the problem requirements:

$$n = \left[\frac{z_\alpha \sqrt{p_0(1-p_0)} + z_\beta \sqrt{p_1(1-p_1)}}{p_1 - p_0} \right]^2$$
$$and\ c = np_0 + z_\alpha \sqrt{np_0(1-p_0)}$$

1. _____

[4] ANSI/ASQC Z1.4. *Sampling Procedures and Tables for Inspection by Attributes*. American Society for Quality Control (1993). Milwaukee.

The probability of rejection for the *double sampling plan S(n,c)* is obtained by substituting the probability of "acceptance" for that of "rejection", in above equations. For such p = 0.09 the new probability of incorrect rejection is 0.018, instead of just P{X < 16} = 1-0.957 = 0.043, which would be the corresponding probability for a fixed sample plan, with n = 20 and c = 16:

$$
\begin{aligned}
&P\{\text{Re } jecting.Drug\} = P\{\text{Re } jecting.Initially\} + P\{Initially.Inconclusive.Then.\text{Re } ject.at.2nd\} \\
&= P\{First.Successes \le 13\} + P\{First.Successes = 14.or.15.and.Combined.Successes < 33\} \\
&= \sum_{x=0}^{13} Bin(x; n = 20,\ p) + Bin(x = 14; n = 20,\ p) \times \{1 - [Bin(x = 19; n = 20,\ p) + Bin(x = 20; n = 20,\ p)]\} \\
&+ Bin(x = 15; n = 20,\ p) \times \{1 - [Bin(x = 18; n = 20,\ p) + Bin(x = 19; n = 20,\ p) + Bin(x = 20; n = 20,\ p)]\} \\
&= 0.0024 + 0.0089 \times (1 - (0.2701 + 0.1215)) + 0.0319 \times (1 - (0.2851 + 0.2701 + 0.1215)) = 0.018
\end{aligned}
$$

Notice how such double sampling scheme, when compared to the fixed sample test, not only (1) increases the probability of accepting a good drug, but (2) reduces the probability of rejecting a good one, even if the initial test results are inconclusive. These are the strongest advantages of multiple stage plans which, in many cases, far out weight their extra cost and effort.

In addition, double (multiple) sampling also reduces long run averages of the sample size. Now, the random variable "sample size" is probabilistic (varies with every case). Its "Expected Value", known as ASN or "Average Sample Number", depends on the real value of the percent "p" of infection, which is the parameter under test.

The *ASN for double sampling* is obtained:

$$ASN = E\{SN\} = \sum_{SN} SN \times P\{SN\} = n_1 \times P(n_1) + (n_1 + n_2) \times P(n_1 + n_2)$$

SN (*sample number*) can only be n_1 or $n_1 + n_2$. Then, P (n_1) is the probability of drawing a "first" sample only, which occurs when arriving to a decision at the first sample (with probability 1 - P $\{c_1 \le Y \le c_2\}$). The probability P $\{n_1 + n_2\}$, of drawing a second sample totaling a size of $n_1 + n_2$, occurs when we had an *inconclusive* outcome from the first sample with P $\{c_1 \le Y \le c_2\}$).

For the double sampling plan *S (n_1=20, n_2=20, c_1=14, c_2=15, c_3=33),* described above, let the true "p", be p = 0.8, and let **Y** be the number of survivals obtained in the first sample, of size n_1=20. Then, the probability of taking "no decision" on the first sample, when p = 0.8, is P ($c_1 \le Y \le c_2$) = P ($14 \le Y \le 15$). This yields an ASN = 25.6, smaller than $n_1 + n_2 = 20+20 = 40$.

Clinical trials could also be designed and implemented as done in industrial statistics. This could help reduce clinical trial testing times. If the initial results were conclusively good or bad, such clinical trials could be terminated early. Alternatively, if the initial results were uncertain, then subsequent testing stages could be implemented, to arrive to a safer conclusion.

4.0 The Binomial Sequential Probability Ratio Tests

In *doubly blind clinical trials*, doctors and their patients do not know who is getting the vaccine and who the placebo. But there is usually *a record in the clinical trial protocol* that can be used to conduct statistical procedures such as Binomial Sequential Probability Ratio Test (BSPRT).

SPRT tests are designed *when results are obtained sequentially one at a time*. *SPRT* can also be *assessed one at a time*. And just as with the double sampling scheme, *a test decision* to accept or reject, can be taken *based upon the results* obtained, at any stage.

Let the above-mentioned second research group with access to the information regarding which patients are actually taking the Covid-19 vaccine analyze the incoming results of such group. If a *patient* that has received the vaccine *falls ill* with the virus, it means that the vaccine has failed to work. We can *assign say value p=0.05 for efficient vaccines*, if they are successful in 95% of cases (i.e. they protect 95% or more of vaccinated individuals, and fail on 5% or less). Likewise, we *assign value p=0.2 for inefficient vaccines*, if they are successful in 80% or less of cases (i.e. if they are able to protect only 80% or less of the vaccinated individuals)

Let the researchers determine a *time T, after virus inoculation into a volunteer*, for the vaccine to work (prevent Covid-19 to develop). If after time T a *volunteer has not fallen ill* with Covid-19, the case is considered *a success* and the patient is removed from the experiment. Alternatively, if *the patient becomes infected* before time T, the case is considered *a failure*. Such accountability will be the task of the second research group that will implement the Binomial SPRT.

For the purpose of BSPRT, *each volunteer will consist of a Bernoulli trial*. If the *virus inoculated* volunteer gets *infected* with Covid-19 before time T, it is *a failure (X=1*, with probability p), and he is removed from the experiment. If by time T the volunteer has not fallen ill with Covid-19, it is a *success (X=0*, with probability 1-p) and volunteer will also be removed from the experiment. The value of said probability p will depend on the BSPRT hypotheses H_0 and H_1.

Let **n** be the number of volunteers in the clinical trial, inoculated but not yet infected, and still within their time T. Let **y** be the cumulative number of infected cases, so far. There are two possible **p** values for the *infection percent*: a preferred value (if say, less than 5% of vaccinated cases get infected) and an unacceptable value (if say, more than 20% become infected).

Assume we implement the BSPRT to test the *hypotheses: H_0: $p_0 \leq 0.05$ and H_1: $p_1 \geq 0.2$.* In such setting, every inoculated patient is observed for say, T=20 days after vaccination, and is thence assumed to be an independent Bernoulli trial, with respective probabilities of success p_i, i = 0, 1. The *cumulative number of successes* "y", out *of "n" trials*, is distributed *Binomial* (n,p_i); i = 0, 1.

Define the *Probability Ratio (PR)* as that of the Binomial distributions, *under H_1 and H_0:*

$$\frac{P\{"y".Successes.UnderH_1\}}{P\{"y".Successes.UnderH_0\}} = \frac{Binomial(y;n,p_1)}{Binomial(y;n,p_0)} = \frac{Kp_1^y(1-p_1)^{n-y}}{Kp_0^y(1-p_0)^{n-y}} = \frac{p_1^y(1-p_1)^{n-y}}{p_0^y(1-p_0)^{n-y}}$$

K, the number of *feasible* ways that one can obtain "y" successes out of the "n" trials, cancels. We then find two values A and B such that, at any stage "n" (i.e. having tested "n" volunteers sequentially, one at a time), and having obtained "y" cumulative infections, said PR fulfills:

$$B < \frac{p_1^y (1-p_1)^{n-y}}{p_0^y (1-p_0)^{n-y}} = \left(\frac{p_1}{p_0}\right)^y \left(\frac{1-p_1}{1-p_0}\right)^{n-y} < A$$

We now define two hypothesis test errors: α (probability of rejecting a vaccine with acceptable rate) and β (probability of accepting a vaccine with a high failure rate). Let α = 0.05 and β = 0.1. Then, we calculate, at every stage, the probability ratio PR:

$$P\{PR > A\} = P\left\{\frac{p_1^y (1-p_1)^{n-y}}{p_0^y (1-p_0)^{n-y}} > A\right\} = \beta; P\{PR < B\} = P\left\{\frac{p_1^y (1-p_1)^{n-y}}{p_0^y (1-p_0)^{n-y}} < B\right\} = 1-\alpha$$

The above equations define the S (B,A), *Sequential Probability Ratio Test* (SPRT), as one that compares PR with values A and B at every stage "n", and decides whether to: (i) accept H_0 if PR < B; (ii) accept H_1 if PR > A; or (iii) continue testing another subject, if B < PR < A.

It can be shown (References 1 and 2) that constants A and B can be approximated by:

$$A \cong \frac{(1-\beta)}{\alpha}; B \cong \frac{\beta}{1-\alpha}$$

For *stage 10 of our example*, we consider the *tenth volunteer* on test (n = 10): we have previously observed "y" successes (say, y = 1 infection), as well as eight volunteers having completed their 20 days without infection. Assume we observe a second infection for such *tenth volunteer*.

For such stage 10 case we calculate the Binomial SPRT PR:

$$\frac{P\{y = 2.Successes.UnderH_1\}}{P\{y = 2.Successes.UnderH_0\}} = \frac{Binomial(2;10, p_1)}{Binomial(2;10, p_0)} = \frac{0.2^2 (1-0.2)^{10-2}}{0.05^2 (1-0.05)^{10-2}} = \frac{6.71E-3}{1.66E-3} \cong 4.04$$

Then, we compare result 4.04 with adequate values for A and B and decide to: (i) stop testing and accept (H_1) that the vaccine infection probability is $p \geq 0.2$, and hence the vaccine doesn't work, if the PR value 4.04 is greater than B; (ii) stop and accept (H_0) that infection probability is $p \leq 0.05$, and hence the vaccine works well, if PR value 4.04 is smaller than A; or (iii) test yet one more volunteer and repeat the process, if 4.04 is between values of A and B.

We can simplify the above process and equations, by taking Logarithms in said PR inequality, which defines the region leading to the continuation of the test. The result produces a linear equation which is a function of the number of successes "y", out of the number of trials (stage) "n" implemented thus far, and which is bounded by the Logarithms of values A and B. The coefficients of "a" and "b" of these equations are functions of SPRT p_i, i = 0, 1.

$$\ln(B) < n\ln\left(\frac{1-p_1}{1-p_0}\right) + y\left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)\right\} < \ln(A)$$

$$\ln(B) < an + by < \ln(A); with: a = \ln\left(\frac{1-p_1}{1-p_0}\right); b = \left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)\right\}$$

In our example: $p_0 = 0.05$, $p_1 = 0.2$, n =10, y = 2 and $\alpha = 0.05$; $\beta = 0.1$. Hence, the SPRT coefficients "a" and "b" can be calculated, and values "A" and "B" can be approximated:

$$a = \ln\left(\frac{1-p_1}{1-p_0}\right) = \ln\frac{1-0.2}{1-0.05} \approx \ln(0.842) = -0.172;$$

$$b = \left\{\ln\left(\frac{p_1}{p_0}\right) - \ln\left(\frac{1-p_1}{1-p_0}\right)\right\} = \ln\frac{0.2}{0.05} - \ln\frac{1-0.2}{1-0.05} = 1.558;$$

$$A \cong \frac{(1-0.1)}{0.05} = 18; \Rightarrow \ln(A) = 2.89;. B \cong \frac{0.1}{1-0.05} = 0.11; \Rightarrow \ln(B) = -2.25$$

For our example, at the SPRT 10th stage (n = 10 trials), with y = 2 successes, we get:

$$\ln(B) = -2.25 < -0.17n + 1.558y < \ln(A) = 2.89$$
$$With.: Z = -0.17 \times 10 + 1.558 \times 6 = 1.39$$

The value of SPRT equation Z (n) $= -0.17n + 1.558y = 1.396$ falls inside the *Continuation Region* (-2.25, 2.89) and therefore, we need to observe another volunteer. The representation of said SPRT tests for the horizontal axis is depicted in Figure 1. The test sequence for the example developed here, is given in Table 1. Decision is taken on stage 15 (3.652>2.89).

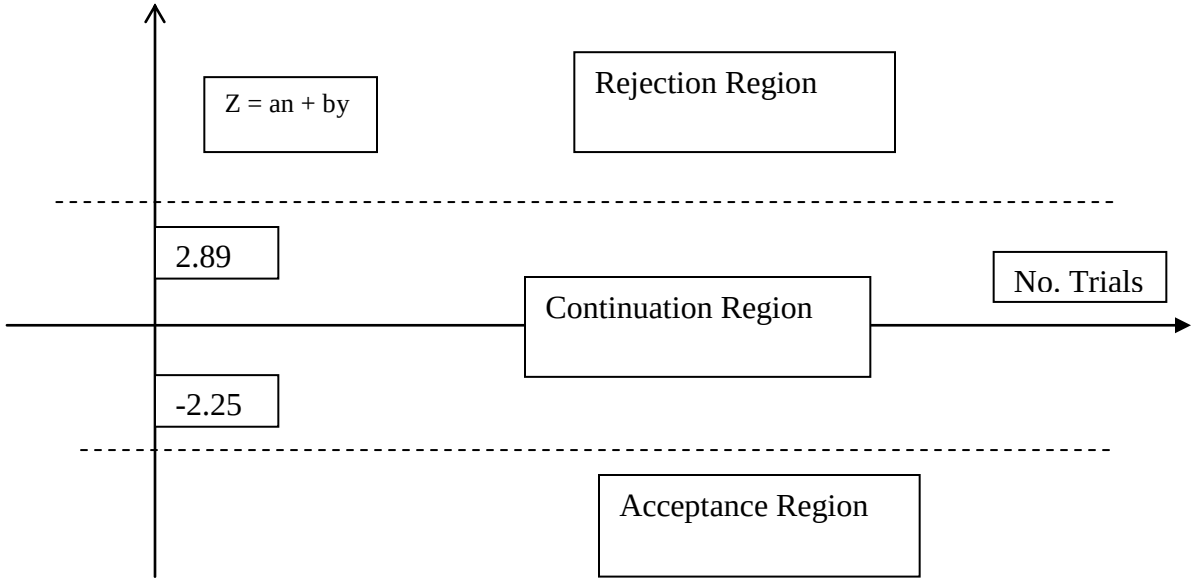

Figure 1: Representation of the SPRT test for horizontal regions.

Table #1: SPRT Evaluation

Stage	CumFail	ZFunction
1	0	-0.172
2	0	-0.344
3	0	-0.516
4	0	-0.688
5	0	-0.860
6	0	-1.032
7	1	0.354
8	1	0.182
9	1	0.010
10	2	1.396
11	2	1.224
12	3	2.610
13	3	2.438
14	3	2.266
15	**4**	**3.652**

The 15th volunteer suffers the 4th cumulative virus infection. The equation value (3.65) is larger than the upper bound (B=2.89). The *SPRT stops* testing and *rejects the hypothesis* H_0: $p_0 \leq 0.05$ that the vaccine protects at least 95%, and instead *assumes the alternative* hypothesis H_1: $p_1 \geq 0.2$ that the vaccine allows an infection rate of at least 20%.

The *SPRT used only n=15 volunteers to discard* this vaccine, which saved time and personnel.

5.0 Discussion

We are struggling with the Coronavarus Pandemic, a new, little researched, and very contagious virus. There is yet no treatment or vaccine against it. To buy us badly needed time to develop them, palliative solutions such as social distancing, testing and contact tracing, isolation of the infected, hand sanitation, hot spot shut-downs etc., are being used. But these are not substitutes.

An (1) *efficient* (protect most people to whom it is administered) and (2) *safe* (do no harm) *vaccine* must be *found*. This implies *trying out multiple vaccines* and then *selecting* the *best ones*. Not all vaccines will work. And, of those who do, some will have troubling characteristics or side effects. *Discovering* these may allow using some vaccines on some cohorts and other vaccines in other cohorts, thus minimizing their negative consequences.

Statistics can help with *the first phases* of *identifying the most effective vaccines* among all the candidates, and *discarding* the inferior ones. The first phases are usually shorter and use fewer volunteers. Thence, the two methods illustrated in this paper are applicable to them.

If an experiment consists of *testing the results of two groups* (vaccine v. placebo) we first need to establish the *sample size* **n** to be used, which depends on the *errors α and β* and the distance $\delta = p_1 - p_0$ defining the difference in vaccine efficiency. In our example $\delta=0.1$ was small, so n=309. If such sample size is too large, and we need a smaller one, we need to give up on *errors α, β, or δ*. Let's make $\delta = 0.3$ instead. The new sample size required would now be only 34.33. The *size of the sample* is then *of great importance*, and we *need methods that reduce* it at the lowest cost.

One of these methods is *double sampling*. The size of the first sample may be too small; the size of both may be too large. Having the possibility of using such *a second sample* only *when* really *needed*, leads to a more efficient way to *reduce the sample size*. Since we will be testing many vaccines with different samples, double sampling will reduce the overall sample size.

Another method discussed is Binomial Sequential Probability Ratio Tests (*BSPRT*). It *requires* that *volunteers be inoculated with the virus (for drugs and treatments this is not a problem, for the patient disease has been diagnosed and treatments either work or not)*. Inoculation allows us, when the vaccine does not work, to (1) be sure volunteers have been infected, and (2) when has infection occurred. These facts allow us to establish thresh-hold T for the virus infection to arise.

Inoculation is not always possible[5]. In such cases, said *volunteers* will need to be released in an environment where they *may become naturally infected*. But then, *we can never be sure* that they did become infected and the vaccine prevented the virus to prosper, *nor the exact date* the virus infection occurred, in order to use thresh-hold T. Doctors and researchers may be able to come up with *some procedure that substitutes* virus *inoculation*, and *provides* similar *information*.

The *second research group* would monitor the trial volunteers inoculated with the virus. Each inoculated volunteer has only two outcomes. One is that *the vaccine works* and, at the end of T pre-assigned days, *the volunteer is virus-free*. The other is that the volunteer becomes infected with the virus, before T days elapse. Such *results* are annotated as they take place (if a tie occurs, break it at random) and are *sequentially incorporated* into the SPRT procedure.

Vaccines that *fail efficiency* tests are *discarded*. Those *that succeed* are passed to *higher levels* (and longer) *clinical trial phases*, where the finalists are selected. But *more important* yet: here the *secondary* (negative) *effects* of the vaccines are *detected*. For example, one efficient vaccine may induce heart problems or strokes in older patients. Another one may induce liver failures on diabetic patients. Such vaccines should not be administered to patients with diabetes, or elderly. But if needed (hundreds of millions of doses will be required) they may be used, accordingly.

Before we close, we want to bring up *two crucial issues*. First, there is a need to *implement virus test*, with samples taken *at random* from the entire population. Such sampling would allow us to *estimate virus prevalence*, something that so far has yet to be done (mostly, *testing for suspected* virus *cases has been done*, so we can confirm, trace and isolate them). The latter approach helps reduce the spread of Covid-19, but *it does not* provide an *estimation* of its *prevalence*.

Secondly, the *current political environment* in the country *does not help*. In December 2019, our attention was concentrated in the Congress accusations against the President. In January 2020, said attention was on the Senate impeachment. During February and March, President Trump minimized the importance of the impending Pandemic. *Valuable preparation time was lost*, that could have reduced deaths. It is *necessary to depoliticize* the struggle against Covid19.

1. ———————————

[5] On August 1900 US Army physician J. Carroll allowed an infection by a mosquito to test the Yellow fever theory. http://www.americaslibrary.gov/jb/progress/jb_progress_yellow_1.html#:~:text=On%20August%2027%2C%201900%2C%20Carroll,mosquitoes%20transmitted%20the%20feared%20disease.

6.0 Conclusions

There is an *urgent need* to find both a *treatment and a vaccine* to combat Covid-19. For that, we need to *perform clinical trials* with vaccine candidates to select *efficient and safe ones*. There is some *pressure to curtail* such clinical trials in order to have these tools *available soon*.

Statistics, the science that studies variation and searches and analyzes patterns (or lack thereof) using the most efficient methods of data collection, can be used to *help shorten clinical trials* based on science and not on other considerations. The *ASA*, as an institution member of *Civil Society, and individual statisticians*, are in a position to *contribute* in this area.

I *encourage* my colleague *statisticians to propose* other *procedures* that can be used toward this end, to develop *examples* and *tutorials* with them, and *share* them with public health researchers. I also encourage *public health scientists and practitioners to try* these methods out, in their work.

We are all *together*, in this battle against the Coronavarus Pandemic.

Bibliography

1. Romeu, J. L. *Understanding Binomial Sequential Testing*. RAC START Vol. 12, No. 2. https://web.cortland.edu/matresearch/BinSeqSTART.pdf
2. Romeu, J. L. Understanding Exponential Sequential Tests. RAC RELTIQUES V. 1, No. 2. https://web.cortland.edu/matresearch/RELTIQUES_V1N2.pdf
3. Romeu, J. L. Determining the Experimental Sample Size. QR&CII Tutorial. Vol. 1, No.1. https://web.cortland.edu/romeu/ExperSampSizeQR&CII.pdf
4. Romeu, J. L. *OC Function and Acceptance Testing*. RAC START. Volume 11, Number 8 https://web.cortland.edu/matresearch/OCCurveSTART.pdf
5. Romeu, J. L. *Censored Data*. RAC START. Volume 11, Number 3. https://web.cortland.edu/matresearch/CensorDatSTART.pdf
6. Duncan, A. J. Quality Control and Industrial Statistics. Fourth Ed. Irwin, Il. 1974.

About the Author

Jorge Luis Romeu retired Emeritus from the State University of New York (SUNY). He was for sixteen years, a Research Professor at Syracuse University, where he is currently an Adjunct Professor of Statistics. Romeu worked for many years as a Senior Research Engineer at the Reliability Analysis Center (RAC), an Air Force Information and Analysis Center operated by IIT Research Institute (IITRI). Romeu received seven Fulbright assignments: in Mexico (3), the Dominican Republic (2), Ecuador, and Colombia. He holds a doctorate in Statistics/O.R., is a C. Stat. Fellow, Royal Statistical Society (RSS), a Member of the American Statistical Association (ASA) and of the American Society for Quality (ASQ). Romeu is a Past ASQ Regional Director, and holds Reliability and Quality ASQ Certifications. He created and directs the *Juarez Lincoln Marti International Ed. Project* (https://web.cortland.edu/matresearch/) which supports i) higher education in Ibero-America and ii) the https://web.cortland.edu/matresearch/QR&CIInstPg.htm

This article is part of our *pro-bono collaboration to the American struggle against Covid-19*, based on our *Proposal for Fighting Covid-19 and its Economic Fallout* that can be read in: https://www.researchgate.net/publication/341282217_A_Proposal_for_Fighting_Covid-19_and_its_Economic_Fallout Such proposal encourages retired professionals like this author to contribute *pro-bono* analyses, based on our long research experience.

His previous work on Covid-19 include a Markov model to study the issue of reopening college: https://www.researchgate.net/publication/343825461_A_Markov_Model_to_Study_College_Re-opening_Under_Covid-19 and a model on the effects of Herd Immunization, https://www.researchgate.net/publication/343345908_A_Markov_Model_to_Study_Covid-19_Herd_Immunization?channel=doi&linkId=5f244905458515b729f78487&showFulltext=true and on some socio-economic and racial issues affected by Covid-19, https://www.researchgate.net/publication/343700072_A_Digression_About_Race_Ethnicity_Class_and_Covid-19 and *A Markov Chain Model for Covid-19 Survival Analysis* https://www.researchgate.net/publication/343021113_A_Markov_Chain_Model_for_Covid-19_Survival_Analysis *An Example of Survival Analysis Applied to Covid-19 Data*, found in https://www.researchgate.net/publication/342583500_An_Example_of_Survival_Analysis_Data_Applied_to_Covid-19, *Multivariate Statistics in the Analysis of Covid-19 Data*, and *More on Applying Multivariate Statistics to Covid-19 Data*, both of which can also be found in: https://www.researchgate.net/publication/341385856_Multivariate_Stats_PC_Discrimination_in_the_Analysis_of_Covid-19, as all above-cited articles, also in our *ResearchGate* web page: https://www.researchgate.net/publication/342154667_More_on_Applying_Principal_Components_Discrimination_Analysis_to_Covid-19 *Design of Experiments to the Assessment of Covid-19*: https://www.researchgate.net/publication/341532612_Example_of_a_DOE_Application_to_Coronavarius_Data_Analysis, Outsource: https://www.researchgate.net/publication/341685776_Off-Shoring_Taxpayers_and_the_Coronavarus_Pandemic and *reliability methods for ICU* units: https://www.researchgate.net/publication/342449617_Example_of_the_Design_and_Operation_of_an_ICU_using_Reliability_Principles