



Statistical Evaluation of Clinical Trials Under COVID-19 Pandemic

Shein-Chung Chow¹ · Wei Zhang^{2,3}

Received: 30 April 2020 / Accepted: 9 June 2020 / Published online: 24 June 2020
© The Drug Information Association, Inc 2020

Abstract

When conducting clinical trials under COVID-19 pandemic, protocol deviations and/or protocol violations are inevitably encountered due to possible environment change which may have an impact on the accuracy and reliability of clinical evaluation of the test treatment under investigation. Protocol deviations and/or violations include, but are not limited to, eligibility criteria, testing procedure, dose and dose regiment, treatment duration and clinical operation of the intended trial. FDA published guidance in March 2020 to assist investigators and institutional boards (IRBs) in assuring the safety of trial participants, maintaining compliance with good clinical practice, and minimizing risks to trial integrity during the COVID-19 pandemic. The purpose of this article is to proposal methods for statistical evaluation in terms of (i) possible shift in target patient population and (ii) assessment of reproducibility of clinical studies conduct under COVID-19 pandemic environment. The proposed methods can be used to determine whether the clinical results of the intended trials are acceptable from both statistical and regulatory perspectives.

Keywords COVID-19 · Shift in patient population · Reproducibility probability

Introduction

COVID-19 is a novel illness that can affect people's respiratory system. It's caused by a virus called coronavirus. Symptoms of COVID-19 include are a cough, fever, and shortness of breath. Thus far, there is no specific antiviral treatment for COVID-19. Current treatment aims to relieve the symptoms until recover. COVID-19 is also known as SARS-CoV-2, nCov, or 2019 Novel Coronavirus. On March 11, as the result of the COVID-19 outbreak that has swept into at least 114 countries, the World Health Organization (WHO) announced COVID-19 is a pandemic viral disease.

As COVID-19 spreads in the United States (US), it is hard to fully understand how suddenly and dramatically, our daily lives have changed. Everyday life including living and working environments continues to change for people across the country. Similarly, COVID-19 has also challenges with the environment for conducting clinical trials. As a result,

on March 13, 2020, President Trump declared a national emergency in response to COVID-19. The US Food and Drug Administration (FDA) recognizes that the pandemic's impacts on the conducting clinical trials. Challenges eventually arise from quarantines, site closures, travel limitations, interruptions to the supply chain for the investigational product, or other considerations if site personnel or trial subjects become infected with COVID-19. On March 18, 2020, FDA published first released guidance on *Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic* [4] to assist industry, investigators, and institutional boards (IRBs) in conducting clinical trials under the COVID-19 pandemic. The guideline provided several updates to support continuity and response efforts to this pandemic. On March 27, 2020, (updated on April 28, 2020), the European Medicines Agency (EMA) published guidance on *Clinical Trial Management During the COVID-19 Pandemic* [3] to advise the sponsor on how they should adjust the management of clinical trials and participants during the COVID-19 pandemic.

FDA guidance is to provide general considerations to assist sponsors in assuring the safety of trial participants, maintaining compliance with good clinical practice (GCP), and minimizing risks to trial integrity during the COVID-19 pandemic. As indicated in the FDA guidance, the impact of COVID-19 may lead to difficulties in

✉ Wei Zhang
michelle.zhang@ppccro.com

¹ Duke University School of Medicine, Durham, NC, USA

² Protech Pharmaservices Corporation, Shanghai, China

³ Potomac, MD, USA

meeting protocol-specified procedures, including administering or using the investigational product or adhering to protocol-mandated visits and laboratory/diagnostic testing. Challenging issues such as increase of safety/risk concern from clinical perspectives and quality, validity, and integrity of the intended trial from statistical perspective are inevitably encountered when conducting clinical trials under the COVID-19 pandemic. The purpose of this article is not only to post these challenging issues, but also to propose methods for evaluation of the validity of clinical trials conducted under the COVID-19 pandemic in terms of (i) the assessment of sensitivity index for testing possible shift in target patient population and (ii) the assessment of reproducibility probability for determining whether the results are reproducible. The proposed methods are helpful in assisting FDA for regulatory decision-making of clinical trials conducted under the COVID-19 pandemic.

The remaining of this article is organized as follows. In the next section, challenging issues that are commonly encountered during the conduct of clinical trials under COVID-19 pandemic are posted from both clinical and statistical perspectives. “[Statistical Evaluation of Clinical Studies](#)” section proposed two useful methods for evaluation of the validity of clinical trials conducted under COVID-19 pandemic. These two methods include the method for the assessment of sensitivity index for testing possible shift in target patient population, and the method for the assessment of reproducibility probability for determining whether the results are reproducible. “[An Example](#)” section gives an example to illustrate the proposed methods for evaluation of the validity of the clinical trials conducted under COVID-19. Some concluding remarks are given in the last section of this article.

Challenging Issues Under COVID-19

One of the major concerns regarding conducting clinical trials under the COVID-19 pandemic is probably that there is a potential shift in target patient population in terms of shift in mean response and inflation in the variability associated with the response. In addition, it is also a great concern that whether the observed clinical results are reproducible if we shall conduct a similar trial under normal clinical environment without COVID-19 pandemic. In what follows, we will examine these concerns from both clinical operation and statistical perspectives.

Clinical Operation Perspectives

The shift in the target patient population could result in increase of safety/risk concern and diminish of efficacy of the

test treatment under investigation. The increase of safety/risk concern could include (i) protocol deviation and/or violation (e.g., eligibility criteria), (ii) patient compliance, and (iii) concurrent illness and medication. These safety/risk concerns may be due to change in clinical trial environment under COVID-19 pandemic. Similarly, the diminish of efficacy could include (i) protocol deviation and/or violation, (ii) delay of dosing and/or shorten of treatment duration, (iii) patient compliance, (iii) concurrent illness and medication, and (iv) possible treatment-by-center interaction. These efficacy concerns may be due to change in clinical trial environment under COVID-19 pandemic.

During the COVID-19 pandemic, these safety/risk and efficacy concerns may arise from quarantines, site closures, travel limitations, interruptions to drug supply of the investigational product, patient adherence and compliance, or site personnel or trial subjects become infected with COVID-19. Consequently, strategy for preventing possible operational biases is necessarily implemented for good clinical practice (GCP) to ensure data quality and integrity of the intended trial.

Statistical Perspectives

When conducting clinical trials under the COVID-19 pandemic, operational biases are evitably encountered. The operational biases could introduce bias and variation in data collection consequently have negative impact on data quality, validity and the integrity of the trial. Most importantly, it may jeopardize the assessment of treatment effect (in terms of accuracy and reliability) of the test treatment under investigation.

In practice, thus, it is suggested that the sources/types of bias and variation due to the change in clinical environment by COVID-19 be identified, eliminated if possible, and controlled whenever possible. Chow and Liu [2] classified the sources/types of bias and variation into the following categories that (i) expected and controllable, (ii) expected but uncontrollable, (iii) unexpected but controllable, and (iv) unexpected and uncontrollable.

Independent Data Monitoring Committee

To overcome the challenging issues raised from both clinical and statistical perspectives, it is strongly recommended that an independent data monitoring committee (IDMC) be established to ensure quality, validity, and integrity of the intended trial under the COVID-19 pandemic for good statistical practice (GSP) and clinical practices (GCP).

Statistical Evaluation of Clinical Studies

Shift in Target Patient Population

Let μ_0 and σ_0 be the expected mean response and the corresponding standard deviation of the response from the intended trial under the clinical environment without COVID-19. Similarly, denote μ_1 and σ_1 be the expected mean response and the corresponding standard deviation of the response from the intended trial under the COVID-19 pandemic environment. Since the patient population of with and without COVID-19 pandemic are similar but different, it is reasonable to assume that $\mu_1 = \mu_0 + \varepsilon$ and $\sigma_1 = C\sigma_0$ ($C > 0$), where ε is referred to as the shift in population mean and C is the inflation factor of the population standard deviation. Thus, the (treatment) effect size E_1 adjusted for standard deviation of the intended trial under COVID-19 pandemic can be expressed as follows:

$$E_1 = \left| \frac{\mu_1}{\sigma_1} \right| = \left| \frac{\mu_0 + \varepsilon}{C\sigma_0} \right| = |\Delta| \left| \frac{\mu_0}{\sigma_0} \right| = |\Delta|E_0, \tag{1}$$

where $\Delta = (1 + \varepsilon/(\mu_0))/C$ and E_0 and E_1 are the effect size of clinically meaningful importance observed from the intended trial with and without COVID-19 pandemic, respectively. Δ is referred to as a sensitivity index measuring the change (shift) in effect size between the patient population with and without COVID-19 pandemic.

As it can be seen from (1), if $\varepsilon = 0$ and $C = 1$, $E_0 = E_1$. That is, the effect sizes of the two populations (with and without COVID-19 pandemic) are identical. In this case, we claim that the results observed from the patient population under the COVID-19 pandemic are consistent with those from the intended trial without COVID-19 pandemic. In other words, there is little impact on the clinical results obtained from the intended trial under COVID-19 as compared to those obtained without COVID-19 pandemic environment. E_1 can be used to support regulatory approval of the treatment under investigation. Note that shift in population mean (i.e., change in ε) could be offset by the inflation/deflation of variability (i.e., change in C). As a result, the sensitivity index may remain unchanged while the target patient population has been shifted. To provide a better understanding, Table 1 provides a summary of the impacts of various scenarios of population shift (ε) and change in variability (C).

As indicated by Chow and Shao [1], in many clinical trials, the effect sizes of the two populations could be linked by baseline demographics or patient characteristics if there is a relationship between the effect sizes and the baseline demographics and/or patient characteristics (e.g., a covariate vector). In practice, however, such covariates may not exist or exist but not observable. In this case, the sensitivity

Table 1. Changes in Sensitivity Index.

$\varepsilon/\mu(\%)$	Inflation of Variability		Deflation of Variability	
	$C(\%)$	Δ	$C(\%)$	Δ
-20	120	0.667	80	1.000
-10	120	0.750	80	1.125
-5	120	0.792	80	1.188
0	120	0.833	80	1.250
5	120	0.875	80	1.313
10	120	0.917	80	1.375
20	120	1.000	80	1.500

index may be assessed by simply replacing ε and C with their corresponding estimates [1]. Intuitively, ε and C can be estimated by

$$\hat{\varepsilon} = \hat{\mu}_1 - \hat{\mu}_0 \text{ and } \hat{C} = \hat{\sigma}_1/\hat{\sigma}_0,$$

respectively. Thus, the sensitivity index can be estimated by

$$\hat{\Delta} = \frac{1 + \hat{\varepsilon}/\hat{\mu}_0}{\hat{C}}. \tag{2}$$

The sensitivity index can be assessed to evaluate whether there is a potential shift in target patient population under COVID-19. Let (Δ_L, Δ_U) be the regulatory acceptable range for Δ . Thus, we claim that there is little impact on COVID-19 if $\hat{\Delta}$ falls within (Δ_L, Δ_U) .

Note that in practice, the shift in population mean (ε) and/or the change in inflation/deflation of population standard deviation (C) could be random. If both ε and C are fixed, the sensitivity index can be assessed based on the sample means and sample variances obtained from the two populations. In real world problems, however, ε and C could be either fixed or random variables. In other words, there are three possible scenarios: (i) the case where ε is random and C is fixed, (ii) the case where ε is fixed and C is random, and (iii) the case where both ε and C are random. These possible scenarios have been studied in Lu et al. [6].

Assessment of Reproducibility

The purpose of reproducibility analysis is to determine whether the current findings of the clinical study under COVID-19 is reproducible. Let H_0 be the null hypothesis that there is no difference in mean response between a test drug product and a control (e.g., placebo). When H_0 is concluded, the test drug product is not considered to be effective. Let T be the test statistic based on the responses from the clinical trial conducted under COVID-19. The trial is considered a positive trial if the observed value of T leads to the rejection of the null hypothesis H_0 . Let H_a be the

alternative hypothesis which states that H_0 is not true. Suppose that the null hypothesis H_0 is rejected if and only if $|T| > c$ based on a two-sided hypothesis, where c is a positive known constant. Under H_a , the probability of observing a positive clinical result is given by

$$P(|T| > c|H_a), \tag{3}$$

which is referred to as the power of the test T .

Suppose now that one clinical trial was conducted and the result is positive. What is the probability that the second trial will produce a positive result, i.e., whether the positive result from the first trial is reproducible? Mathematically, because the two trials are independent, the probability of observing a positive result in the second trial when H_a is true is still given by (3) regardless of whether the result from the first trial is positive or not. However, information from the first clinical trial should be useful in the evaluation of the probability of observing a positive result in the second trial. This leads to the concept of reproducibility probability, which is different from the power defined by (3).

In general, the reproducibility probability is a person’s subjective probability of observing a positive clinical result from the second trial, when a positive result from the first trial was observed. For example, the reproducibility probability can be defined as the probability in (3) with all unknown parameters replaced by their estimates from the first trial (Goodman [5]). In other words, the reproducibility probability is defined to be an estimated power of the second trial using the data from the first trial. Perhaps a more sensible definition of reproducibility probability can be obtained using the Bayesian approach (see, e.g., Shao and Chow [7]). Under the Bayesian approach, the unknown parameter under H_a , denoted by y is a random vector with a prior distribution $p(y)$ assumed to be known. Thus

$$P(|T| > c|H_a) = P(|T| > c|\theta)$$

and the reproducibility probability can be defined as the conditional probability of $|T| > c$ in the second trial, given the data set x observed from the first trial, i.e.,

$$P(|T| > c|x) = \int P(|T| > c|\theta)\pi(\theta|x)d\theta$$

where $T = T(y)$ is based on the data set y from the second trial and $\pi(\theta|x)$ is the posterior density of y , given x .

To determine whether the current clinical results under COVID-19 are reproducible, we first find (P_L, P_U) , where P_L and P_U are the reproducibility probability calculated based on the confidence upper bound and lower bound of the observed variability associated with the response, respectively. If P_L is greater than a pre-specified number P_0 , then we claim that there is evidence that the findings obtained from the current trial is reproducible.

Proposal for Regulatory Decision-Making

For evaluation of the intended clinical trial, the criteria for (i) possible shift in patient population and (ii) assessment of reproducibility as described in the previous subsections can be used to determine the validity and integrity of results obtained under the COVID-19. The proposed criteria are summarized in Table 2.

Based on Table 2, recommendations regarding whether the results obtained from the intended trial under COVID-19 is acceptable or need to be adjusted are summarized in Table 3.

As it can be seen from Table 3, the trial conducted under COVID-19 is considered a success provided that (i) there is no evidence of shift in patient population and (ii) the results are reproducible (scenario 1). On the other hand, the trial fails if there is evidence of shift in patient population and the observed results are not reproducible (scenario 4). In the case where (i) there is no shift in patient population and (ii) results are not reproducible (scenario 2), it is suggested that the significance level α should be adjusted for achieving a desirable (empirical) power (i.e., reproducibility probability). For illustration purpose, Table 4 lists reproducibility probabilities \hat{P} with different values of $T(x)$ under a two-group parallel design with and without equal variances.

In the case where (i) there is a shift in patient population and (ii) results are reproducible (scenario 3), treatment effect should be adjusted for the shift for a more accurate and reliable assessment of the treatment effect.

As it can be seen from Table 4, if p value is less than 0.01, the reproducibility probability is greater than 90% (i.e., $\hat{P} = 0.91$). This suggests that the decrease of significance level will increase the probability of reproducibility.

An Example

For illustration purpose, consider an example concerning the evaluation of the efficacy of ONGENTYS for the adjunctive treatment to levodopa/carbidopa in patients with Parkinson’s disease (PD) experiencing “off” episodes (ONGENTYS 2020). The efficacy of ONGENTYS was evaluated in two Phase 3, multiregional, double-blind, randomized, parallel-group, placebo-controlled trials, BIPARK-1 and BIPARK-2.

Table 2. Criteria for Evaluation of Clinical Trials Under COVID-19.

		Results are Reproducible	
		Yes	No
Shift in patient population	Yes	Yes/yes	Yes/no
	No	No/yes	No/no

Table 3. Proposed Criteria for Regulatory Decision-Making.

Scenario	Status	Description	Regulatory Decision
1	No/yes	There is no shift in patient population Results are reproducible	Success
2	No/no	There is no shift in patient population Results are not reproducible	May adjust for α
3	Yes/yes	There is shift in patient population Results are reproducible	May adjust for shift
4	Yes/no	There is shift in patient population Results are not reproducible	Fail

Table 4. Reproducibility Probabilities \hat{P} .

T(x)	Known σ^2		Unknown σ^2 ($n = 30$)	
	p value	\hat{P}	p value	\hat{P}
1.96	0.050	0.500	0.060	0.473
2.05	0.040	0.536	0.050	0.508
2.17	0.030	0.583	0.039	0.554
2.33	0.020	0.644	0.027	0.614
2.58	0.010	0.732	0.015	0.702
2.81	0.005	0.802	0.009	0.774
3.30	0.001	0.910	0.003	0.890

In BIPARK-2, patients ($n = 427$) were randomized to treatment with either one of two doses of ONGENTYS once daily ($n = 283$) or placebo ($n = 144$). The intention to treat (ITT) study population included patients treated with ONGENTYS 50 mg once daily ($n = 147$) or placebo ($n = 135$). The primary efficacy endpoint was the change in mean absolute OFF-time based on 24-h patient diaries completed 3 days prior to each of the scheduled visits. ONGENTYS 50 mg significantly reduced mean absolute OFF-time compared to placebo. Table 5 provides absolute OFF-time (hours) change from baseline to endpoint of the BIPARK-2 study (Source: https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/212489s000lbl.pdf; ONGENTYS[®]).

We first assess whether there is a possible shift in target population. Suppose that $(\Delta_L, \Delta_U) = (90\%, 110\%)$ is considered acceptable range for sensitivity index by the

regulatory agency. As it can be seen from Table 5, the absolute OFF-time change from baseline for the treatment group (ONGENTYS) and placebo are given by -1.98 and -1.07 , respectively. Suppose there is a shift (diminishing) of mean change from baseline in treatment group by 0.198 , i.e., $\frac{\epsilon}{\mu} = \frac{-0.198}{-1.980} = 0.1$ or 10% and there is 20% inflation of variability due to clinical trial environment change as a result of COVID-19. By Table 1, these shifts lead to a sensitivity index (Δ) of 0.750 (i.e., $\Delta = 75\%$), which lower than the lower bound of the acceptable range, i.e., $\Delta_L = 90\%$. The diminishing in efficacy by 0.198 h is an indication that there is a shift in target patient population. The shift in target patient population, which is non-negligible in regulatory decision-making, may be due to change in clinical trial environment as result of COVID-19.

For assessment of reproducibility probability of the findings of the study, assuming that the observed mean change from baseline for the treatment group and the placebo group are given by -1.98 and -1.07 , respectively, are the true mean change from baseline, there is a more than 95% probability of reproducibility if the study shall be conducted under similar experimental conditions with a total of 282 subjects (147 subjects in the treatment group and 135 subjects in the placebo group) assuming that the standard deviation of the observed difference in mean change from baseline is 0.31 , which is obtained as $[-0.287 - (-1.523)]/4$. The high reproducibility probability is expected due to the relatively small p value (p value = 0.008) observed from the study.

Table 5. BIPARK-2—Absolute OFF-time (Hours) Change from Baseline to Endpoint.

	N	Baseline Mean (SE)	LS Mean Change from Baseline (SE)	Placebo-subtracted Difference (95 % CI)	Adjusted p Value ^a
Placebo	135	6.12 (0.200)	-1.07 (0.239)	–	
ONGENTYS 50 mg	147	6.32 (0.183)	-1.98 (0.230)	-0.91 ($-1.523, -0.287$)	$p = 0.008$

CI confidence interval, LS least squares, N total number of patients, SE standard error.

^aAdjusted p values were calculated using Dunnett's alpha level adjustment to control for multiplicity.

One of the limitations for the proposed methods for ensuring validity of trial results is that shift in population mean may be offset by the inflation/deflation of shift in scale parameter. This is mainly because the derived sensitivity index is an aggregated measurement combining both the shift in location parameter and shift in scale parameter.

Remarks

For another example, on April 6, 2020, Immunomedics announced their ASCENT study to be stopped for efficacy while the estimated study completion date was about to read out in mid-2020. ASCENT is an international, multi-center, open-label, randomized, phase 3 study in patients with metastatic Triple-Negative Breast Cancer refractory or relapsing after at least 2 prior chemotherapies (including a taxane) for their metastatic disease. Eligible patients were randomized 1:1 to receive either sacituzumab govitecan or treatment of physician's choice. The primary endpoint for the study is progression-free survival, and secondary endpoints include overall survival and objective response rate, duration of response among others. This decision of stopping the trial was based on the recommendation by the independent DSMC, with no pre-specified interim analysis, during its recent routine review of the ASCENT study. Stopping the trial early would help assuage the potential dropout and censoring problems from COVID-19. However, it is suggested the proposed methods for assessment of (i) possible shift in target patient population and (ii) reproducibility probability be performed to determine whether the assessment of treatment effect has been contaminated due to the change in clinical trial environment as the result of COVID-19 pandemic.

Concluding Remarks

COVID-19 has changed living and working environment since WHO announced COVID-19 is a pandemic viral disease in March. The outbreak of COVID-19 has also changed the environment for conducting clinical trials especially global or multiregional clinical trials due to travel limitations. FDA published guidance to assist sponsors, investigators and institution review boards (IRBs) by providing general considerations for conducting clinical trials under COVID-19 pandemic. Some challenging issues such as concerns due to increase of safety/risk and diminishing efficacy (from clinical perspectives) and quality, validity, and integrity of the intended trial (from statistical perspectives) raised. It is then of interest to determine whether the clinical trials conducted under COVID-19 pandemic can provide

accurate and reliable assessment of the efficacy and safety of the test treatment under investigation.

In this article, two useful methods for evaluation of the validity of clinical trials conducted under the COVID-19 are proposed. One method regarding the assessment of sensitivity index is to test whether the target patient population has been shifted due to the change in clinical trial environment as result of COVID-19 pandemic. The other method regarding the assessment of reproducibility probability is to determine whether the results obtained from the current trial under the COVID-19 are reproducible if a similar trial is to be conducted under the same experimental conditions without COVID-19 pandemic. The trial is considered a successful trial if it passes the test for sensitivity index (i.e., there is no evidence of shift in target patient population) and there is evidence that the results are reproducible, while the trial is considered failed if it fails to pass the test for sensitivity index and the results are not reproducible. When there is no evidence of shift in target patient population but the results are not reproducible, it is suggested the α be adjusted for achieving a desired reproducibility probability but at the same time maintaining statistical significance for regulatory approval.

References

1. Chow SC, Shao J. Inference for clinical trials with some protocol amendments. *J Biopharm Stat.* 2005;15:659–66.
2. Chow SC, Liu JP. *Design and Analysis of Clinical Trials.* 3rd ed. New York: Wiley; 2013.
3. EMA. *Guidance on the Management of Clinical Trials During the COVID-19 (Coronavirus) Pandemic.* Amsterdam: The European Medicines Agency; 2020.
4. FDA. *Guidance for Industry, Investigators, and Institutional Review Boards - FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency.* Silver Spring: The United States Food and Drug Administration; 2020.
5. Goodman SN. A comment on replication, p-values and evidence. *Stat Med.* 1992;11:875–9.
6. Lu Y, Kong YY, Chow SC. Analysis of sensitivity index for assessing generalizability in clinical research. *Jacobs J Biostat.* 2017;2(1):009.
7. Shao J, Chow SC. Reproducibility probability in clinical trials. *Stat Med.* 2002;21:1727–42.
8. The BIPARK-2 Study, A Phase 3, Multiregional, Double-Blind, Randomized, Parallel-Group, Placebo-Controlled Trial for Evaluation of the Efficacy of ONGENTYS for the Adjunctive Treatment to Levodopa/Carbidopa in Patients with Parkinson's Disease (PD) Experiencing "OFF" Episodes. ONGENTYS®. https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/212489s000lbl.pdf.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.