General Psychiatry

Review of current controversial issues in clinical trials

Shein-Chung Chow,[1] Susan S Chow,[2] Annpey Pong[3]

[1]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA
[2]Kansas City University, College of Osteopathic Medicine, Kansas City, Missouri, USA
[3]Merck & Co Inc, Kenilworth, New Jersey, USA

Correspondence to
Susan S Chow;
susan.chow@kansascity.edu

ABSTRACT

In clinical development, adequate and well-controlled randomised clinical trials are usually conducted to evaluate the safety and efficacy of test treatment under investigation. The purpose is to ensure that there is an accurate and reliable assessment of test treatment under study. In practice, however, some controversial issues inevitably appear despite the compliance of good clinical practice. These debatable issues include, but are not limited to, (1) appropriateness of hypotheses for clinical investigation, (2) feasibility of power calculation for sample size requirement, (3) integrity of randomisation/blinding, (4) strategy for clinical endpoint selection, (5) demonstrating effectiveness or ineffectiveness, (6) impact of protocol amendments and (7) independence of independent data monitoring committee. In this article, these controversial issues are discussed. The impact of these issues in evaluating the safety and efficacy of the test treatment under investigation is also assessed. Recommendations regarding possible resolutions to these issues are provided whenever possible.

INTRODUCTION

In clinical development, adequate and well-controlled randomised clinical trials are often conducted for evaluation of the safety and efficacy of a test treatment under study. The purpose is to provide substantial evidence in support of regulatory submission for demonstration of the safety and efficacy of the test treatment under investigation. To provide an accurate and reliable assessment of the test treatment under investigation, adequate and well-controlled randomised clinical trials should follow good clinical practice (GCP) at different phases (ie, phases 1 to 3) of clinical development. In practice, some controversial statistical issues inevitably occur regardless of the compliance of GCP. Controversial issues, referred to as debatable issues, are commonly encountered during the conduct of clinical trials. These debatable issues could be raised from (1) compromises between theoretical and real-world practices, (2) miscommunication, misunderstanding and/or misinterpretation in medical and/or statistical perception among regulatory agencies, clinical scientists and biostatisticians and (3) disagreement, inconsistency, miscommunication/misunderstanding and errors in clinical practice.

Basically, these controversial issues present conceptually different perspectives from clinicians (investigators/sponsors), biostatisticians and regulatory reviewers for evaluation of the test treatment under investigation. The major concern from the clinicians is whether the observed difference is of clinical meaning; whereas, the biostatisticians are interested in determining whether the observed difference is of any statistical meaning (ie, whether the observed difference is by chance alone). On the contrary, the reviewers from regulatory agencies, such as the United States Food and Drug Administration (FDA), would like to make sure if the observed clinically meaningful difference (clinical benefits) has reached statistical significance before they can approve the test treatment under investigation. Most regulatory reviewers consider these debatable issues as review issues (in the sense that only regulatory reviewers can make the final judgement). Thus, a clinical trial is considered successful in clinical development when it meets the expectations of clinicians, biostatisticians and regulatory reviewers.

In the subsequent sections, background and relevant information regarding these controversial issues and their potential impacts on clinical development are briefly described. Whenever possible, resolutions for each controversial issue are provided. A final concluding remark is given in the last section.

APPROPRIATENESS OF HYPOTHESES FOR CLINICAL EVALUATION

For clinical evaluation of the safety and efficacy of a test treatment under investigation, a traditional approach is to first test for a null hypothesis that there is no treatment difference in efficacy under a valid trial design. If there is sufficient power to correctly detect a clinically meaningful difference (or treatment effect) showing that such a difference truly exists, we are then in favour of the alternative hypothesis that the test treatment is

Table 1 Examples of drug products being withdrawn

Drug name	Usage	Company	Reasons	Years on market
Accutane	Acne	Hoffman-LaRoche	Birth defects, etc	1982–2009
Baycol	Cholesterol reduction	Bayer AG	Kidney failure	1998–2001
Bexlra	Pain relief	GD Searle	Cardiovascular serious adverse event	2001–2005
Cylert	Attention deficit hyperactivity disorder/Attention deficit disorder	Abbott Labs	Liver toxicity	1975–2010
Duract	Pain killer	Wyeth-Ayerst	Liver damage	1997–1998
Vioxx	Pain relief	Merck	Heart attack and stroke	1999–2004
Zelmid	Anti-depressant	AstraZeneca	Higher risk of suicide	1982–1982

efficacious. The test treatment will be approved by the regulatory agency, such as the United States FDA, if there appears no evidence of safety concerns.

However, this typical approach based on efficacy alone could lead to the withdrawal of drug products owing to safety concerns after approval. As an example, table 1 provides a list of drug products being withdrawn, which were marketed between 1982 and 2010.

As an alternative to the traditional approach, Chow and Shao[1] suggested testing composite hypotheses to take into account safety and efficacy (see table 2). In practice, common approaches for clinical evaluation of efficacy are testing for hypotheses of superiority (S), non-inferiority (N) and (therapeutic) equivalence (E). To assess for safety, the investigator typically examines the safety profile relative to the adverse events and other safety parameters to ascertain if the test treatment is better (superiority), at least as good as (non-inferiority) or similar (equivalence) to the control treatment.

As an example, an investigator may be interested in testing non-inferiority in efficacy and superiority in safety of a test treatment compared with a control. In this case, we can consider testing the null hypothesis that H_0: *not NS*, where N denotes the non-inferiority in efficacy and S represents superiority of safety. We would reject the null hypothesis and be in favour of the alternative hypothesis that H_a: *NS*, that is, the efficacy of the test treatment is non-inferior to the control group and the safety of the test treatment is superior to the control group. For testing the null hypothesis of H_0: *not NS*, the test statistics is derived based on the null hypothesis. In addition, the derived

test statistics can then be evaluated to achieve the desired power under the alternative hypothesis and the required sample size in the trial can be estimated by power analysis. The objectives of selecting the required sample sizes in the intended trial are for (1) demonstrating the non-inferiority of the test treatment compared with the control group in efficacy and (2) exhibiting the superiority of the safety profile of the test treatment compared with the control group at a prespecified level of significance. Note that the significance levels for efficacy and safety need to be chosen to control the overall type I error rate at the α level (such as 5%). A sample size increase would be expected after switching from a single hypothesis testing to a composite hypothesis testing.

Feasibility of power calculation for sample size requirement

In clinical trials, a power analysis for sample size calculation (power calculation) is often performed to select a sample size required for achieving the study objective with a desired power of correctly detecting a clinically meaningful difference (or treatment effect) at a prespecified level of significance. A required sample size is typically determined based on an appropriate statistical test derived under the null hypothesis and a valid study design. The test statistic is then evaluated under the alternative hypothesis to achieve the desired power at a prespecified level of significance. In cases of clinical studies with extremely low incidence rates or rare disease clinical trials, power calculation for sample size may not be feasible.

For illustration purpose, consider an example concerning a safety study of a test treatment for treating patients with diabetes. A pharmaceutical company is asked to conduct a diabetes study with an extremely low incidence rate of glycosylated haemoglobin (HbA1c) to demonstrate that there is no safety concern of the test treatment under investigation. However, the incidence rate of HbA1c is extremely low, which is 3 per 10 000.

FDA indicates that 1 per 10 000 is of clinical importance and suggests that a clinical safety study should be powered to detect a 1 per 10 000 (clinically meaningful difference) at the 5% level of significance, assuming that

Table 2 Composite hypotheses for clinical evaluation of both safety and efficacy

		Safety		
		S	N	E
Efficacy	S	SS	SN	SE
	N	NS	NN	NE
	E	ES	EN	EE

E, equivalence; N, non-inferiority; S, superiority.

the true incidence rate is 3 per 10 000. A power analysis is then performed to determine the sample size required for achieving the study objective with a desired power of 80% at the 5% level of significance. The result of power calculation indicates that a total sample size of 784 684 is required to have a desired power of 80% for correctly detecting a 1 per 10 000 incidence rate of HbA1c at the 5% level of significance. In this case, the power calculation is definitely not feasible for FDA-suggested clinical safety study.

Alternatively, the sponsor considered a probability statement to justify a selected, more feasible sample size of 800. In other words, assuming that the incidence rate of HbA1c is 3 per 10 000 (ie, 0.000 3), with a selected sample size of 800, we expected to not observe a single event during the conduct of the study. FDA eventually accepted the approach and issued an approvable letter based on the reason that there is not sufficient evidence to demonstrate that the test treatment is not safe. It should be noted that the term 'safe' is not the same as the term 'not unsafe'. Sample size justification based on probability statement allows us to demonstrate that the test treatment is 'not unsafe' rather than 'safe' with a relatively small sample size.

INTEGRITY OF RANDOMISATION/BLINDING

In clinical trials, randomisation and blinding are essential for minimising the bias that may be owing to subjective selection bias and prior knowledge of the treatment codes. For this purpose, block randomisation is often considered to prevent treatment imbalance, especially when the sample size is small or subjects' characteristics change overtime. In practice, however, treatment imbalance may still occur, especially in multicentre clinical trials. Breaching blindness in clinical trials is another serious problem leading to subjective judgement and observational bias. In practice, the questions listed below are commonly asked:

1. How do we test the integrity of blinding in clinical trials?
2. In a comparative trial, what is the difference in the probability of correctly guessing the treatment code for a blocking size 2 compared with that of the blocking size 4?

Regarding the first question, one method to determine whether the blindness is seriously violated is to ask patients to guess their treatment codes during the study or at the conclusion of the trial prior to unblinding. In some cases, investigators may also be asked to guess patients' treatment codes. Once the guesses are recorded on the case report forms and entered into the database, the integrity of blinding can be tested (Chow and Shao).[2] For illustration purpose, consider the example described in Karlowski *et al*.[3] To evaluate the difference between the prophylactic and therapeutic effects of ascorbic acid for common cold, a double-blind placebo-controlled study was conducted by the National Institutes of Health

Table 3 Results of patients' guess

| Patient's guess | Actual treatment assignment | |
	Ascorbic acid	Placebo
Ascorbic acid	40	11
Placebo	12	39
Do not know	49	39
Total	101	89

(NIH). On completion of the study, a questionnaire was distributed to each of the 190 subjects enrolled in the study for them to guess which treatment they received. Results from the subjects who completed the study are summarised in table 3.

Consider a single site parallel design comparing $a \geq 2$ treatments. Let A_{ij} be the event that a patient in the jth treatment group guesses that he/she is in the ith group, $i=1,\ldots,a,a+1$, where $i=a+1$ defines the event that a patient does not guess (or answers 'do not know'). To test the integrity of blinding, Chow and Shao[2] considered testing the following null hypothesis:

H_0: $P(A_{ij})=P(A_{1j})$ for any i and j

If the null hypothesis holds, then the blindness is considered to be preserved. Note that the hypothesis H_0 can be tested using the Pearson's χ^2 test. Based on data given in table 3, the observed Pearson's χ^2 statistic is 31.3. Thus, the null hypothesis of independence is rejected at a significance level of <0.001 (ie, the p value is smaller than 0.001). Thus, the blindness is not preserved and, hence, the integrity of blinding is in doubt.

To address the second question, Wang and Chow[4] considered two fixed block randomisation design with block size of 4 and 2 when comparing a placebo or an active control and a test treatment under investigation. In addition, Wang and Chow also examined the probabilities of observing various degrees of imbalance and investigated the probabilities of correctly guessing the treatment under three types of prior knowledge.[4] The results showed that smaller block size design is more likely to maintain treatment balance than larger block size; however, the difference of imbalance between the two designs decreases when the sample size gets larger. The number of subjects per site in multicentre trials also has impact on the degree of imbalance. Additionally, both block size and prior knowledge have impact on the probability of guessing the treatment right. Large sample size in conjunction with varying block size is suggested for keeping treatment balance and avoiding the chance of guessers to guess the treatment right.

For illustration purpose, probabilities of guessing treatment codes right for a small clinical trial are given in table 4.

One of the controversial issues regarding the randomisation/blinding is whether a formal statistical test for the integrity of the randomisation/blinding should be performed at the end of the clinical trial (especially when

Table 4	Probability of correctly guessing treatment codes		
Blocking size	**N=4**	**N=8**	**N=16**
2	0.2500	0.0625	0.0039
4	0.1667	0.0278	0.0008

Note that analysis on investigators' guesses of patients' treatment codes can be performed similarly.

significantly positive results are observed). In addition, what action should be taken if a positive clinical trial fails to pass the test for the integrity of the randomisation/blinding? Regarding the impact of different blocking sizes in the randomisation of a clinical trial, it should be noted that the knowledge of the blocking size may increase the probability of guessing the treatment codes right for the investigator. Although the increase of the blocking size may decrease the probability of guessing the treatment codes right, it will also increase the probability of mixing up the randomisation schedule and the possibility of treatment imbalance at the end of the trial.

STRATEGY FOR CLINICAL ENDPOINT SELECTION

In psychiatry clinical trials, commonly considered clinical endpoints, there are rating scales for functional outcomes that include, but are not limited to, (1) depression, (2) cognitive functioning, (3) motor functioning and (4) proximity to diagnosis. Suppose that these four rating scales are available for psychiatry clinical studies. These four rating scales will lead to (1) four single primary endpoints, (2) six co-primary endpoints, (3) four triad endpoints and (4) one composite endpoint. Thus, there are a total of 15 endpoints for measuring functional outcomes.

In practice, it is usually not clear which of the 15 study endpoints can best inform the disease status and/or measure the treatment effect. Moreover, although different study endpoints may not translate one another, they may be highly correlated to one another. In clinical trials, it should be noted that different endpoints may result in different sample sizes. Thus, it is critical to select a promising endpoint that can achieve the study objective of the intended study with a desired power at a prespecified level of significance. This, however, is a very controversial issue (or purely a review issue as pointed out by regulatory agencies) in regulatory submissions.

To address this controversial issue, Filozof *et al*[5] proposed using a utility function to develop a therapeutic index by fully utilising all clinical information collected from all available study endpoints. The idea is briefly described below. Let e_i be the ith endpoint, where $i = 1, 2, \ldots, 15$. Consider a therapeutic index (TI) in the following:

$$TI = f\left(w_i e_i, w_2 e_2, \ldots, w_{15} e_{15}\right),$$

where f is a utility function and w_i is the weight of e_i depending on the level of evidence (eg, p value) observed from e_i. Compared with a given endpoint (say e_i for some i), the therapeutic index has the following good statistical

properties (Chow and Huang).[6] First, the therapeutic index can also detect the true treatment effect provided that the given ith endpoint has successfully detected the treatment effect. Second, if the therapeutic index has successfully detected the treatment effect, the ith given endpoint may not.

DEMONSTRATING EFFECTIVENESS OR NOT INEFFECTIVENESS

For approval of a new drug product, the sponsor is required to provide substantial evidence regarding the safety and efficacy of the drug product under investigation. A typical approach is to conduct adequate and placebo-controlled clinical studies and test the following point hypotheses:

$$H_0 : \text{ineffectiveness versus } H_a : \text{effectiveness} \quad (1)$$

The rejection of the null hypothesis of ineffectiveness is in favour of the alternative hypothesis of effectiveness. Most researchers interpret the rejection of the null hypothesis as demonstration of the effectiveness of the alternative hypothesis. It should be noted, however, that 'in favour of effectiveness' does not imply 'the demonstration of effectiveness'. Alternatively, Chow and Huang[6] indicated that the alternative hypotheses of (1) should be 'not H_0: not ineffectiveness' rather than 'H_a: effectiveness'. In other words, H_a=not H_0 as follows:

$$H_0 : \text{ineffectiveness versus } H_a \text{ (not } H_0) : \text{not ineffectiveness} \quad (2)$$

As it can be seen from H_a in (1) and (2), the concept of 'effectiveness' and the concept of 'not ineffectiveness' are not the same. Not ineffectiveness does not imply effectiveness. Thus, the traditional approach for clinical evaluation of the drug product under investigation can only demonstrate 'not ineffectiveness' but not 'effectiveness'. In practice, we typically test the null hypothesis at the α=5% level of significance. However, many researchers prefer testing the null hypothesis at the α=1%. If the observed p value falls between 1% and 5%, we claim the test result is 'inconclusive'. In placebo-controlled studies, conceptually, 'not ineffectiveness' includes the portion of 'inconclusiveness' and 'effectiveness' (figure 1).

As indicated in Chow,[7] the concept of demonstrating 'not ineffectiveness' rather than demonstrating 'effectiveness' is useful in support of regulatory submission of
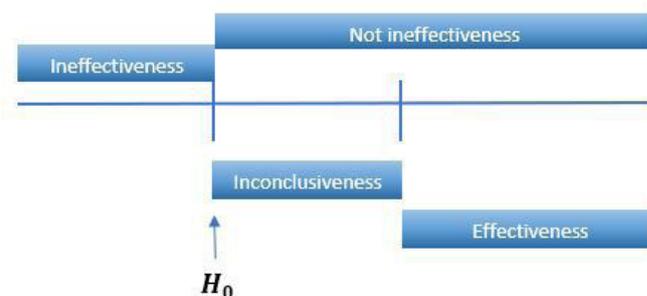

Figure 1 The relationship between 'effectiveness' and 'not ineffectiveness' in placebo-controlled studies.

rare diseases drug development and can be used in drug development with normal conditions.

Impact of protocol amendments

During the conduct of a clinical trial, it is not uncommon to have 2–5 protocol amendments. These amendments are necessary to describe the changes that have been made and the rationales or justifications (both statistically and clinically) behind the changes to ensure the quality, integrity and validity of the intended clinical trial. If the changes are major, the originally target patient population under study may have shifted to a similar but different target patient population. In this case, the original clinical trial may have become a totally different trial, which cannot address the scientific/medical questions that the clinical trial is intended to answer. In practice, it is suggested that potential risks for introducing additional bias/variation as a result of protocol amendments should be carefully evaluated. It is important to identify, control and hopefully eliminate/minimise the sources of bias/variation owing to protocol amendments.

One of the major impacts of many protocol amendments is that the target patient population may have been shifted during the process, especially when significant changes or modifications are made to inclusion/exclusion criteria of the study. Thus, one of the controversial issues in this regard is whether the conclusion drawn (by ignoring the population shift) at the end of the trial is accurate and reliable.

Denote by (μ, σ) the target patient population. After a given protocol amendment, the resultant (actual) patient population may have been shifted to (μ_1, σ_1), where $(\mu_1 = \mu + \varepsilon)$ is the population mean of the primary study endpoint and $(\sigma_1 = C\sigma)$ $(C>0)$ is the population standard deviation (SD) of the primary study endpoint. The shift in target patient population can be characterised as follows:

$$E_i = \left| \frac{\mu_1}{\sigma_1} \right| = \left| \frac{\mu + \varepsilon}{C\sigma} \right| = \left| \Delta \right| \left| \frac{\mu}{\sigma} \right| = \left| \Delta \right| E,$$

where $\Delta = (1 + \varepsilon/\mu)/C$, and E and E_1 are the effect sizes before and after population shift (or protocol amendment), respectively. Chow et al[8] and Chow and Chang[9] refer to Δ as a sensitivity index measuring the change in effect size between the actual patient population and the original target patient population. Chow and Chang[9] indicated that the impact of protocol amendments on statistical inference owing to shift in target patient population can be studied through a model that links the moving population means with some covariates (see also Chow and Shao[10]). In many cases, however, such covariates may not exist or exist but not be observable. In this case, Chow et al[11] suggested that inference on Δ be considered to measure the degree of shift in location and scale of patient population based on a mixture distribution by assuming that the location or scale parameter is random.

In summary, frequent protocol amendments could result in a moving target patient population, which makes the intended trial more difficult (if not impossible)

to address the medical or scientific questions that the study intends to answer. In practice, however, there are no existing regulations regarding how many protocol amendments after the initiation of a clinical trial can be issued. Thus, it is strongly suggested that regulatory guideline/guidance regarding (1) levels of changes and (2) number of protocol amendments that are allowed should be developed in order to maintain the quality, integrity and validity of the intended study.

INDEPENDENCE OF DATA MONITORING COMMITTEE

In clinical trials, especially during the late phase of clinical studies, an independent data monitoring committee (IDMC) is often established to maintain the integrity and validity of the intended clinical trials. The IDMC is responsible for conducting ongoing safety monitoring and/or performing interim analyses for efficacy. The IDMC supposes to be independent of any activities related to clinical operation (project team) of the study. The role and responsibility are often clearly stated in the Charter of the IDMC. An IDMC typically comprises experienced physicians and statisticians and, if established, has a separate IDMC staff support to perform responsibilities more efficiently. A typical organisational flow for an IDMC is given in figure 2.

As indicated in figure 2, the primary responsibilities of the IDMC include ongoing data safety monitoring and possibly interim analysis for early stopping owing to safety, futility and/or efficacy. The IDMC will make a recommendation to the sponsor regardless of whether the sponsor chooses to accept or reject the recommendation. The good intention of the IDMC will ensure the quality, integrity and validity of the clinical trial.

One of the most controversial issues when utilising an IDMC is probably the independence of the IDMC. In addition, the following controversial issues have been raised:

1. If there is any wrongdoing in the conduct of the intended clinical trial, should the IDMC be encouraged to communicate to regulatory agencies?
2. Should additional burden be placed on the IDMC if adaptive design methods are utilised?

These controversial issues have an impact on the quality, integrity, validity and success of clinical trials conducted throughout various phases of clinical development.

In practice, however, some sponsors will make every attempt to direct the function and activity of IDMC. In some cases, they have been successful, and in many cases, they have failed. The following is a summary of commonly seen issues at IDMCs of clinical trials across therapeutic areas.

First, the sponsor usually will draft an IDMC charter that outline the role/responsibility and function/activity of an IDMC before consulting the IDMC members. This prevents IDMC members from having the chance to review the charter until the first meeting. Note that the initial IDMC meeting is usually a teleconference call
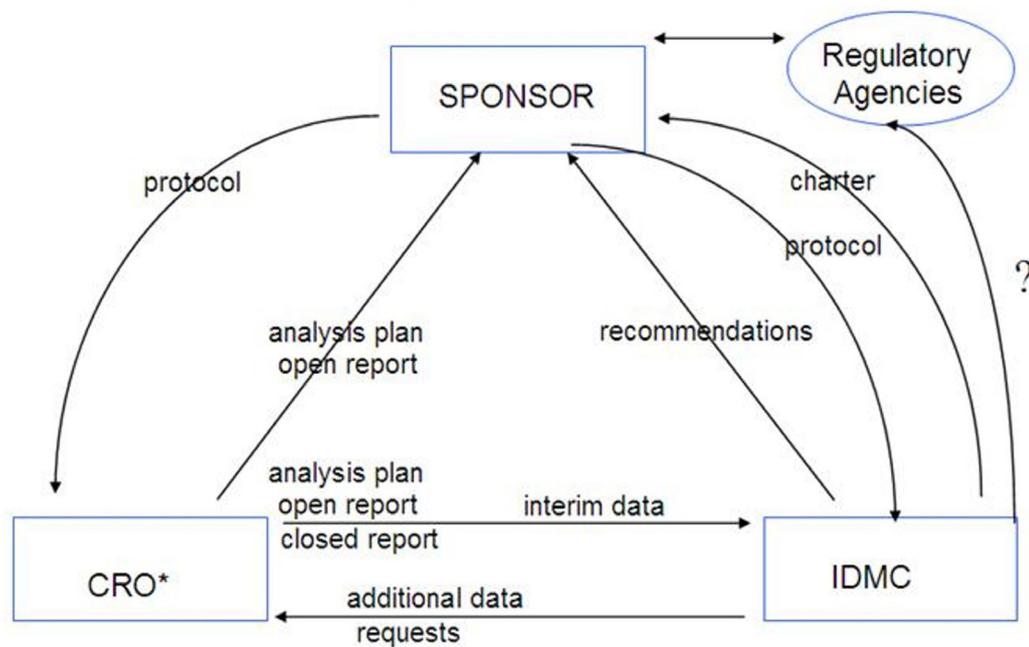
Figure 2 Organisational flowchart for IDMC. IDMC, independent data monitoring committee; CRO, contract research organization.

rather than a face-to-face in order to save on costs. Since IDMC members are usually key opinion leaders in the subject area, they may not have the chance to thoroughly review the charter prior to the meeting. Consequently, the charter is approved in a hurry. In some cases, the sponsor may have begun to enrol patients prior to the initial IDMC meeting. Although this is definitely not a GCP, it does occur.

Second, in some cases, some IDMC members may have strong opinions regarding the design and analysis of the study protocol and/or charter. In this case, the sponsor should communicate with these IDMC members rather than replace them. For creditability, the regulatory agencies, such as the United States FDA, recommend documentation for the reasons to replace IDMC members. However, the sponsors are not in compliance to this suggestion when the IDMC members are replaced prior to the official establishment of the IDMC. Another issue is related to protocol amendments. In many cases, the sponsor may have changed/issued protocol amendments or modified randomisation schedules without consulting with IDMC members. This has brought difficulties for the IDMC to perform their job responsibilities.■

CONCLUDING REMARKS

In addition to the controversial issues described in previous sections, there are many other controversial issues that are commonly encountered in the conduct of clinical trials. Just to name a few, these controversial issues include, but are not limited to, (1) validation of an instrument (or a questionnaire) for subjective evaluation of a test treatment under investigation (Did we ask the right questions?), (2) strategy for handling of missing value imputation (Can we use a legal procedure, that is, statistical method for missing data imputation, to illegally make up clinical data for analysis?), (3) determination of non-inferiority margin in active control trials (What if there is a disagreement between the principal investigators or sponsors and regulatory reviewers?), (4) the issue of recording replicates in QT/QTc interval prolongation for cardiotoxicity (Is a recording replicate really a replicate?), (5) power calculation for multi-regional clinical trials (Is the selected sample size at specific regions sufficient?), (6) the use of simulation in clinical trials (Clinical trial simulation is 'a' solution not 'the' solution), (7) the issue of multiplicity (Should we always adjust for alpha for multiple comparison?), (8) clinical studies utilising complex innovative design such as adaptive trial design (Can the type I error rate be controlled?), (9) the modernisation of traditional Chinese medicine (TCM) (Can current regulatory requirements for Western medicine be applied to TCM directly?) and (10) the assessment of interchangeability for biosimilar products (Is it possible to show that there is same therapeutic effect in any given patient according to FDA guidance?)

Controversial issues are commonly encountered during the conduct of clinical trials. Thus, accuracy and reliability are concerns when assessing the treatment effect at the end of the clinical trials. Shao and Chow[12] recommended evaluating the probability of

reproducibility as a tool to monitor the performance of an approved drug product.[12] The reproducibility probability is the chance of observing significant (positive) results of future clinical trials that are to be conducted under similar circumstances (ie, experimental conditions) given the observed clinical results. Note that if p value is slightly less than 5%, which is a prespecified level of significance, the reproducibility probability could be less than 50% or less than a tossing of an unbiased coin. Thus, evaluating the reproducibility probability can be deemed useful, especially after postapproval by both the sponsor and the regulatory agencies to protect patients from an unexpected risk of test treatment.

REFERENCES

1. Chow SC, Shao J. *Statistics in drug research*. New York, NY: Marcel Dekker, Inc, 2002.
2. Chow S-C, Shao J. Analysis of clinical data with breached blindness. *Stat Med* 2004;23:1185–93.
3. Karlowski TR, Chalmers TC, Frenkel LD, *et al*. Ascorbic acid for the common cold. a prophylactic and therapeutic trial. *JAMA* 1975;231:1038–42.
4. Wang P, Chow SC. The impact of block size on treatment imbalance and blinding integrity in randomized clinical trials. *Biostatistics and Biometrics* 2021.
5. Filozof C, Chow S-C, Dimick-Santos L, *et al*. Clinical endpoints and adaptive clinical trials in precirrhotic nonalcoholic steatohepatitis: facilitating development approaches for an emerging epidemic. *Hepatol Commun* 2017;1:577–85.
6. Chow S-C, Huang Z. Demonstrating effectiveness or demonstrating not ineffectiveness - A potential solution for rare disease drug product development? *J Biopharm Stat* 2019;29:897–907.
7. Chow S-C. Innovative thinking for rare diseases drug development. *AJBSR* 2020;7:292–6.
8. Chow S-C, Shao J, Hu OY-P. Assessing sensitivity and similarity in bridging studies. *J Biopharm Stat* 2002;12:385–400.
9. Chow SC, Chang M. *Adaptive design methods in clinical trials*. New York, NY: Chapman and Hall/CRC Press, Taylor & Francis, 2006.
10. Chow S-C, Shao J. Inference for clinical trials with some protocol amendments. *J Biopharm Stat* 2005;15:659–66.
11. Chow S-C, Chang M, Pong A. Statistical consideration of adaptive methods in clinical development. *J Biopharm Stat* 2005;15:575–91.
12. Shao J, Chow S-C. Reproducibility probability in clinical trials. *Stat Med* 2002;21:1727–42.

Shein-Chung Chow, PhD, is a Professor at the Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA. Dr Chow is also a special government employee (SGE) appointed by the FDA as an Oncologic Drug Advisory Committee (ODAC) voting member and Statistical Advisor to the FDA. Dr Chow was the Editor-in-Chief of the Journal of Biopharmaceutical Statistics (from 1992 to 2020) and the Editor-in-Chief of the Biostatistics Book Series at Chapman and Hall/CRC Press of Taylor & Francis Group. He was elected Fellow of the American Statistical Association in 1995 and was an elected member of the ISI (International Statistical Institute) in 1999. Dr Chow is the author or co-author of over 310 methodology papers and 32 books including Design and Analysis of Clinical Trials, Adaptive Design Methods in Clinical Trials, and, most recently, Innovative Methods for Rare Diseases Drug Development.