

Risks in Finding Doses for a New Drug

NAITEE TING

Abstract

The biggest risk in new drug development is either unaware of, or under-estimate the potential risks in designing clinical trials. Among all challenges in drug development, the most critical one is about finding the appropriate dose(s) for the study drug in treating patients. Designing dose finding clinical trials involves in many potential risks. In practice, most of the expensive failures in drug development originated not from “We did not know”, rather, the mistake is “We thought we knew”. In other words, greatest risks came from lack of awareness of underlying assumptions. This manuscript attempts to discuss some of these risks and make recommendations to reduce risks in the design of dose finding clinical trials. This is not a complete list of risks, but it is only starting the discussion.

1. BACKGROUND

New drug research and development starts with a molecule discovered by laboratory scientists. If the drug company (the sponsor) develops this molecule believes that it can potentially be a successful drug in treating certain type of chronic conditions or diseases, then the sponsor identifies this molecule as a drug candidate, and invest resources and capitals to develop this candidate into a new drug. The drug development process includes both pre-clinical development and clinical development. In pre-clinical development, this drug candidate goes through pharmacology, toxicology, and animal pharmacokinetics/pharmacodynamics (PK/PD) tests. The drug substance needs to be formulated into drug products for animal, and eventually for human testing. Only after the candidate successfully passed these pre-clinical tests, it can then be progressed into clinical development.

The major difference between clinical trials and non-clinical experiments is that clinical trials are scientific tests applied to live human bodies. Pre-clinical development includes *in vitro*, *in vivo* experiments, chemical development, as well as other tests, all are outside of live human body. In clinical development of new drugs, the drug candidate goes through Phases I, II, and III clinical testing. If successful, then all of the study results obtained from non-clinical and clinical testing are documented into a submission package known as a new drug application (NDA), or a biologic license application (BLA). This NDA or BLA is then submitted to the regulatory agency (e.g., Food and Drug Administration – FDA, in the U.S.) for approval. If a new drug is approved by FDA, then Phase IV clinical studies are initiated to assess long term and additional real world experiences of the drug’s impact on human beings [2].

If a molecule can eventually be efficacious to treat diseases, that means this molecule has to change the human

physiological system. If it does not change the human system, it would simply be a placebo. If it changes the human biology, then the molecule could be toxic. Hence it is thought that “all drugs are toxic”. If this is the case, how come there are so many drugs approved by FDA and available in the retail pharmacy stores? The answer is that for every approved drug, there is a non-toxic dose. In other words, for each drug, if the dose is too high, it would be toxic, if the dose is too low, the drug does not work. From this point of view, every step of the new drug development process can be thought of as studying the dose range of the drug candidate. From a clinical development point of view, the upper limit of a dose range is known as the maximally tolerable dose (MTD). The lower limit is considered as the minimally efficacious dose (MinED). For every drug, only the doses between MinED and MTD can be considered as efficacious and safe.

Given this understanding, the four phases of clinical development in new drugs can be realized as a process about finding the appropriate dose or doses for this drug candidate. In Phase I, it is the first time for the candidate is exposed to live human being. The development team has to be very cautious to ensure that human subjects participating in Phase I clinical trials are not exposed to high toxic risks. It is natural that the starting dose of first Phase I trial should be low. Typical Phase I trials are dose escalation trials. That means, trial participants are exposed to very low doses first. If there is no observed adverse events (AE), then a next group of participants are recruited to test at a higher dose. If no AE, then another group of participants is tested at an even higher dose. This process would continue until the MTD is identified.

Another important objective of Phase I clinical development is to study the PK/PD characteristics of the drug candidate in human subjects. Pharmacokinetics (PK) is to learn about what does the body do to the drug, and pharmacodynamics (PD) is about what does the drug do to the

body. In PK, the scientists study the human absorption, distribution, metabolism, and elimination (ADME) of the drug candidate. In order to have a better understanding of these PK properties, participants recruited in Phase I clinical trials tend to be healthy normal volunteers, instead of patients. The main reason is that the PK properties observed in patients could have been compromised by their disease(s). Another important reason is that patients usually take other medications to treat their condition, or disease. These other medications may interact with the drug candidate under study, and hence the PK data obtained from patients may not reflect the true ADME in human.

Therefore, in traditional drug development for the treatment of chronic diseases, Phase I clinical data are mostly obtained from healthy normal volunteers. Clinical efficacy may not be observed from Phase I clinical trial participants because they do not have the disease under study. Accordingly, Phase II trials are designed to recruit patients with the target disease the drug candidate is developed to treat. In other words, before Phase II, all of the pre-clinical experiments and the Phase I clinical data cannot be evaluated for clinical efficacy from patients with the disease under study. Thus the first Phase II study tends to be a proof of concept (PoC) clinical trial. If the concept is proven (test drug is efficacious in treating the disease), then the sponsor invest more resources to further develop it. If not, stop further development.

A typical first Phase II PoC study randomizes patients into two treatment groups – test drug and placebo [6]. These participants are with the disease or condition the study drug is indicated to treat. The PoC trial is designed with a statistical hypothesis test. Sample size is calculated using the Type I and Type II error rates, together with the assumed treatment difference, and within group variance. The sample size reflects the risks and assumptions associated with this primary statistical hypothesis. This study is known as PoC because if the candidate fails to demonstrate clinical efficacy, then further development of this candidate will be stopped. Results of the PoC study will guide the sponsor to make a “Go/NoGo” decision on this molecule. In this design, the test drug dose is usually selected as the MTD obtained from Phase I. The reason is that if the highest allowable dose of the drug candidate cannot demonstrate clinical efficacy when compared against placebo, then the concept is not proven and a “NoGo” decision can easily be made. There will be no further development of this test drug.

After this study, if the PoC results are good and the decision is “Go”. Then the next Phase II clinical trial would be a dose ranging trial which include multiple test doses against a placebo control. A typical dose ranging trial includes four treatment groups – placebo, low dose, medium dose, and high dose of the drug candidate. It is hoped that this dose ranging trial will help find the MinED for the indication this molecule was developed to treat. The objective of Phase II clinical development is first to check PoC – if the

candidate does not deliver clinical efficacy, then stop development; and the second objective is that if the decision is “Go”, then Phase II results should guide the Phase III dose selection. One or a few of the doses will be tested in long-term, large-scale Phase III trials in order to establish drug efficacy and safety. Based on the Phase III results, for each tested dose, if it is efficacious and relatively safe, then that dose can be considered “Approvable” by regulatory agencies.

Therefore, in a few sentences to describe the entire clinical development process can be that –new drug development is about finding the appropriate dose or doses for this molecule under development. If the dose is too high, the drug will be toxic. If the dose is too low, it does not work. In Phase I clinical trials, doses are tested from low to high. The Phase I deliverables are PK/PD and MTD. Then in Phase II, the two objectives are PoC and dose ranging – with a hope to lower test doses from MTD to MinED. If both Phases I and II are successful, then a dose range for the drug candidate can be framed between MinED and MTD. Given this knowledge, one or a few doses are selected for Phase III long-term, large-scale trials to confirm. Suppose Phase III results demonstrate the tested doses are efficacious and safe, then it should be approvable from the regulatory point of view.

Drug development is an expensive, high risk business. Throughout the entire new drug development process, the most important question is about finding the right range of doses to treat patients. This manuscript points out the risks associated with designing clinical trials for dose finding purposes. Section 2 introduces the unpredictability of human behavior, Section 3 lays out that sample sizes are always limited in any clinical trial, especially in Phase II. Section 4 attempts to clarify the two types of clinical questions – confirmatory vs exploratory. Section 5 points out that statistical concepts are difficult to communicate with team members without statistical training. Section 6 states that project team members do not pay sufficient attention to the risk of inconclusiveness. Section 7 clarifies that in designing clinical trials, one of the highest risks is applying models without fully understanding the assumptions behind each model. Given these risks, Section 8 discusses what to consider in designing clinical trials with a hope to reduce some of these risks. Finally Section 9 delivers some concluding remarks.

2. HUMAN BEHAVIOR

Randomized, controlled, double-blind clinical trials are one of the most important scientific breakthroughs in 20th century to make the new drug development and drug approval possible. In a clinical trial, live human subjects are randomized into treatment groups. For the typical placebo-controlled trials, responses from patients receiving the test drug treatment are compared against responses from patients receiving placebo treatment using a statistical hypothesis testing procedure. If the resulted p-value is less

than the pre-specified Type I error rate (α), then the drug developer claims statistical significance, and submit these findings to regulatory agencies (e.g., FDA, in U.S.) for approval. Under this framework, FDA does not approve the drug by saying that the test drug is efficacious, but the drug is approved because “the probability that this test drug is not efficacious is controlled under α ”. Of course, for drug approval, in addition to statistical significance, FDA also requires clinically meaningful treatment benefit.

A clinical trial can also be thought of as a type of scientific experiment. One way to view a scientific experiment could be “taking observations under controlled conditions”. For example, in physics, the double-slit experiment is performed to demonstrate that light and matter satisfy both wave form and particle form. These observations are taken under very well controlled conditions. However, clinical trials are different from these scientific experiments in at least two ways – first, the experiment unit is live human being; and second, there is only a very limited control condition that can be applied to a clinical trial – randomization. Other than randomization, a clinical trial is very similar to a prospective observational study in epidemiology. In most of these studies, epidemiologists observe exposure and diseases with an attempt to assess the association between them.

In a clinical trial, the experiment unit is live human being. Every living person has her/his own free will. A clinical protocol is designed to collect data from live human beings, by live human beings. Given the fact that every individual has their own free will, how can the study team reasonably assume every person conducts the clinical trial, and every person participates in the trial will follow the protocol religiously? In our experience of running clinical trials, we have never seen a protocol being completed without any issue – patients did not meet inclusion/exclusion criteria; individuals are randomized to a wrong block; patients took wrong medication; the same subject entered the same clinical trial for more than once; patients took prohibited rescue drugs, etc. The fact that clinical trial team members, investigators, staff within the clinics, as well as trial participants are all human beings itself can be thought of as a potential risk factor – because human errors are unavoidable.

When designing a clinical trial, it is very important to keep this fact in mind – a clinical trial is to collect data from live human beings by live human beings – anything might go wrong could eventually go wrong. With this understanding, the study team should design robust clinical trials. In other words, even if many people (either trial participants, or people engaged in the conduct of the trial) may not follow the protocol, most of the collected clinical data can still be good enough to help the team making correct decisions. On this basis, the most robust trial design should be the simplest design.

In fact, the assumption that every person will follow the protocol is not realistic. The problems are not only that every person has her/his free will. Even everyone tries to

follow the protocol as close as possible, there can still be misunderstanding, miscommunication, or mistakes. Hence the key principles in designing clinical trials should be simple and robust. Simple designs allow flexibility of trial team members and trial participants. Sometimes they may have different priorities, simple and robust design can still help reach the primary conclusion (most likely, the Go/NoGo decision), even with missing visits, data errors, and protocol violations.

3. SAMPLE SIZE IS LIMITED

Statistical hypothesis testing is based on a sample taken from an underlying population. The size of this sample is associated with probabilities of making wrong decisions (Type I error implies the decision to continue develop a placebo, Type II error means the sponsor gives up a potentially very good drug). This decision, from a regulatory point of view, it is about whether to approve a drug. From the sponsor point of view, it is about whether to continue develop this candidate. In Phase II, it is the “Go/NoGo” decision. Note that for every drug candidate at every stage of development, this “Go/NoGo” decision has to be made. In other words, after completing each experiment, or each clinical trial, the sponsor needs to ask “Should the company spend more investment to develop this candidate?” In reality, there should be an unknown truth that either this drug is truly efficacious, or it does not work. If the truth is the drug does not work, and the null hypothesis is accepted, then the decision is correct. If the truth is the drug works, and the null hypothesis is rejected, it is also a correct decision. However, if the truth is the drug does not work and the null is rejected, then the sponsor or the regulatory agency is making a wrong decision. Statistical hypothesis testing attempts to control the probability of making a wrong “Go” decision, or a wrong approval of a new drug to be under α .

In the design of a clinical trial, sample size calculation is based on α , power, and assumptions about the treatment effect as well as variability. In order to reduce α , and/or to increase power, the study team needs to increase sample size (n). That means, when n is increased, the risk of making a wrong decision can be reduced. However, at early Phase II, the sponsor is not sure whether the drug is efficacious or not, it would be neither ethical, nor cost effective, to recruit a large number of patients for this clinical trial.

In dose ranging trials, many test doses are compared against the placebo control. The objective is to find out test dose or doses that can deliver the anticipated drug efficacy. The idea of “MinED” can be thought of as the study dose such that “among those doses demonstrated drug efficacy, the lowest one”. Or it can be “doses lower than MinED do not deliver drug efficacy”. Note that neither of the above statement can be easily established in a Phase II clinical trial. The reason is that sample sizes in Phase II trials are limited. With small sample sizes, it is not easy to detect statistical significance.

A high level summary of Phase I and II clinical trials can be that – Phase I trials escalate doses from low to high, and Phase II studies lower doses from high to low. The development of any drug is about finding the appropriate dose range. The upper limit of this dose range is MTD, and the lower limit is MinED. In order to find MTD, the key consideration is drug safety, and the key factor for finding MinED is clinical efficacy. In the development of drugs treating chronic diseases, this understanding is well established – when escalating doses, the focus is drug safety – hence there is no need to recruit patients with the target disease. Because clinical efficacy is not the deciding factor in escalating test doses. Similarly, in Phase II, the focus is about efficacy. If the drug candidate does not work, there is no need to further develop this molecule. After the concept is proven, the main interest is to lower doses down to MinED. At this time, the focus would be clinical efficacy associated with the drug candidate. Safety is a secondary consideration in Phase II – without efficacy, further development will be stopped and there is no need to worry about safety. Only after the clinical efficacy is established, next to consider whether the safety profile can be acceptable under the given efficacy.

However, at early Phase II and before PoC results, there is not strong evidence that the test drug can deliver clinical efficacy. Hence it may not be ethical to recruit a large number of trial participants. Also, from the sponsor point of view, it would not be appropriate to invest a huge amount of resources on this drug candidate. Therefore, sample sizes in Phase II clinical studies are limited. When sample size is small, the study results tend to be lack of precision – it is not easy to differentiate the signal from the noise. Accordingly, a clear decision is not easy to make. Typical Phase II studies tend to be designed with limited sample size, which leads to a higher risk of making a wrong decision about whether to continue develop this drug candidate.

Dose ranging trials usually include many test doses, comparing against a placebo control. Traditionally, statisticians use multiple comparison procedures to control the Type I error rate α , when this is the case, sample sizes can be largely increased. Even with newer methodology like MCP_Mod [4], or OLCT [8], the required sample sizes can still be high. Attempting to make decisions, or to make dose recommendations for Phase III, relatively larger sample sizes can still be preferred. One problem with small sample size is the lack of precision. When the total number of patients for a given clinical trial is limited, often the signal is hiding behind the noise, which makes the drug developer unclear regarding the next steps of clinical development. Limited sample size is a real risk in designing, and conducting clinical trials, as well as analyzing or interpreting data from clinical trial results. Small sample size makes the picture ambiguous, and a “Go/NoGo” decision to be difficult.

4. UNDERSTANDING THE QUESTION – IS IT CONFIRMATORY OR EXPLORATORY?

From drug approval point of view, if a new drug at a particular dose or doses is efficacious and safe, then this dose or these doses of the study drug could be approvable. Therefore, from drug development point of view, it is about finding the appropriate range of doses such that with these doses, the test drug is efficacious and safe. At each step in drug development, the question should be about whether the drug candidate demonstrate the expected efficacy, and that it is relatively safe. Because most drug candidates failed to meet both criteria (efficacy and safety), they are “weeded out” during the development process. However, drug efficacy is a confirmatory question and needs to be answered first, especially in Phase II stage. If there is no efficacy, no matter how safe this drug is, there is no need to engage in further development. After all, a placebo is the most safe drug.

Clinical trial designs and clinical development programs are mostly based on statistical thinking and concepts. One example is the statistical hypothesis testing—the null hypothesis is that there is no difference between the drug candidate and the placebo control. However, unless there is strong enough evidence to demonstrate that the test product is significantly different from placebo, scientists would not reject the null hypothesis (that the test treatment is not different from the control treatment). When a decision is made that the test product is efficacious, then it is considered that the probability the drug candidate does not work is controlled under α . This statistical thinking process establishes the foundation of designing an individual clinical trial. Furthermore, statistical thinking can also help guide the planning of an entire clinical development program consists of many clinical trials.

In clinical trials, the only confirmatory question is about this “Go/NoGo” decision based on the primary efficacy endpoint. Any questions other than the primary efficacy question are exploratory in nature. For example, drug safety, subgroup analysis, exploratory analyses, dose-response relationship, ... etc. At the study design stage, sample size calculation is based on the primary hypothesis test of the primary efficacy variable. Alpha protection only applies to this statistical comparison. In some complicated designs, multiple comparison procedures (MCP) are applied to test multiple hypotheses. When this is the case, only those pre-specified comparisons are associated with this experiment-wise error. Any other analysis is not alpha protected. Hence they are considered as exploratory.

In Phase II clinical development, the two most important questions are PoC and dose ranging. In this Phase, the PoC question is a confirmatory question – to Go, or Not to Go. The PoC question is about whether to continue invest and develop this drug candidate. On the other hand, questions for dose ranging trials are exploratory in nature. The objective is to explore the efficacy dose-response relationship

across doses being studied in this trial. This is the point that confused many trial design team members. When the team designs a Phase II dose ranging trial, they are used to ask “which dose or doses deliver the desired treatment efficacy?” From this framework of thinking process, people tend to consider this is a confirmatory question – for a dose to work, that dose should deliver clinical efficacy, if it does not, then consider this is an ineffective dose. But in reality, Phase II dose ranging is exploratory – use Phase II results to recommend dose or doses for Phase III confirmation. Phase II PoC is about “Go/NoGo” (confirmatory), Phase II dose ranging is exploratory, and Phase III is confirmatory.

Recommendation of doses for Phase III clinical trial design should not only be based on drug efficacy. At this point, drug safety is also an important concern. However, when designing Phase II dose ranging trials, the consideration is mostly about drug efficacy. At the design stage, dose ranging is exploratory for Phase II clinical trials. After dose ranging trial results are ready, the team evaluates the efficacy and safety profile of each studied dose, and then make recommendations for Phase III studies. The dose or doses employed in Phase III clinical trial is for confirmatory purposes – if a dose, or a range of doses included in Phase III demonstrate both efficacy and safety, then this dose, or this range of doses are considered approvable by regulatory agencies. In US, typically before engage in Phase III study design, sponsors have an opportunity to meet with FDA and discuss Phase II results, as well as Phase III study designs. This is known as the “End of Phase II Meeting” between sponsor and FDA.

The fact that clinical trial design team does not understand that the nature of clinical questions (confirmatory, or exploratory) is a commonly observed risk. Often times when the team members engage in clinical trial design are not clear about the priority of clinical questions at hand. They can be easily distracted by the complexity of a clinical trial and lost in the details. In fact, when discussing a “Go/NoGo” decision, people can easily be side-tracked by less relevant or less important objectives. In the evaluation of drug efficacy and safety, there can be many measurements, and for each measurement, there can be a variety of factors affect it. This complexity in human body and medical practice leads clinical trial team members to lose focus on the critical confirmatory questions. It is only after the “Go/NoGo” question can be clearly understood by the team, then every other question can simply be thought of as exploratory. In most of clinical trials, many of the questions are sequential – the “Go/NoGo” question needs to be first addressed, before contemplating any other questions. Also, these other questions can also be prioritized in a sequential way to answer.

5. COMMUNICATIONS

Most of the biostatistical applications can be thought of as support for public health. The term “public health” im-

plies that the two most important sciences/professions involved in public health are statistics and medicine. “Public” means the consideration covers a population point of view, not necessarily just individual health. This is the subject that statisticians are trained to address. The profession to study “Health” is, of course, medicine. Therefore, in order to engage in public health activities, statistics and medicine professionals need to communicate, and to work together. However, in reality, this communication can be difficult. In other words, many physicians had difficulties understanding the thinking process from statistical professionals, and some statisticians may misunderstand the nature of the problems, or questions under public health studies.

In drug development and drug regulations, there are at least three sets of very important communications – 1. The communications between sponsor statisticians, and sponsor non-statisticians; 2. The communications between regulatory agency statisticians and agency non-statisticians; and 3. The communications between sponsor and agency. None of these communications is easy. One of the communication gaps came from the training of statisticians and the training of physicians. A Ph.D. in statistics or in biostatistics went through years or decades of rigorous training in the understanding of the mathematical and statistical theories underpin statistical thinking and statistical philosophy. One important part of this training is an abstract way of thinking process – abstract from the real world phenomena into a mathematical model, and this model building process has to be logically rigorous.

One possible cause of this communication gap may have a deep root. Many good statisticians tend to have an introvert personality. They are logic, deep thinkers. They think carefully before they voice their opinion. This fact could lead them to lose opportunities to speak up in meetings. A lot of good statisticians are math majors in their undergraduate training, and they study statistics in graduate school. For their education, the typical classes are that the teacher or the professor lectures, students taking notes. Homework problems are mostly to derive equations or to prove theorems. All of these experiences can be achieved by individual learning, without working with other students.

On the other hand, students with business majors engage in a lot of group discussions, case studies, and team projects. These learning experiences help business students to communicate with other students on a regular basis. After graduation, students with business major (or social science major) are fluent with oral/written communications and group discussions. Students with science majors including physics, chemistry, biology or other physical sciences have to take lab courses. In a lab, students work together to achieve an experiment. This environment provides opportunities for students to communicate with each other in order to make progress. Therefore, scientists with their laboratory experiences are still better communicators than statisticians. Accordingly, many good statisticians tend to be introvert in nature, and

their education does not provide an environment for them to learn to communicate, or to collaborate with others. After receiving a Ph.D. degree in statistics or biostatistics, they join the pharmaceutical industry. In this industry, close communications among team members is critical in progressing any project. These statisticians are in a disadvantageous position because they are lack of training in communicating with non-statisticians.

On the other hand, physicians received medical education for many years, and they usually have practiced in clinics or hospitals, before joining the pharmaceutical industry, or the regulatory agency. In medical school, the objective of training is about patient care. That is, for a patient in front of a physician, medical training is to prepare this physician to make the best diagnosis, and to prescribe the best treatment, in order to help this patient. In other words, the emphasis is about patient care – more appropriately, it is about individual patient care. In public health, the consideration is about health or medical conditions of a group of individuals or a population. This is different from individual care because variabilities across human beings can be huge. Hence the medical training of individual care may not be sufficient to solve public health related problems without the help from statisticians.

Statisticians are trained to make inferences about a population, based on samples collected from this population. In the scientific/professional area of public health, statisticians and physicians need to work together. Clear communications between these two scientific fields become the key to the success. Therefore, an efficient statistician should be able to articulate statistical concepts to non-statisticians unambiguously. This collaboration becomes a mutual education process – the physicians educate the statisticians about the medical conditions, and the statisticians explain the statistical concepts and methods to the physicians. The close collaborations between medicine and statistics are critical across all applications of public health – including drug development, clinical trials, drug regulation, . . . , in addition to epidemiology and observational studies.

To non-statisticians, statistics are abstract, dull, and difficult to understand. Therefore, in communicating statistical ideas with non-statisticians, the statistician has to be clear, precise, and patient. This is why a good statistician/biostatistician in the pharmaceutical industry or in regulatory agencies has to learn and grow on a continuous fashion. This is also why experience becomes an important merit in this profession. Good communication skills are important for a statistician not only when they communicate with physicians. It is also critical when the statistician explains things to the scientists, engineers, team members, and upper management. The learning of communication skills never stops. However, without experiences or training in communications, statistical concepts can be confusing to non-statisticians, and the collaboration becomes less than ideal. Sometimes the non-statisticians can be confused.

6. INCONCLUSIVENESS

In practice, most drugs failed during the development process. In the drug discovery group, scientists innovate many new molecules as drug candidates each year. These candidates are progressed into the pre-clinical development, including pharmacology, toxicology, animal pharmacokinetics/pharmacodynamics (PK/PD), chemical development, and formulation. Only the best candidates can pass these stringent testing and enter the clinical development. In other words, many of the molecules discovered may not be able to pass the pre-clinical selection process. Then, during the clinical Phase I and II stages, more drug candidates failed and stopped for further development. Finally, out of tens, hundreds, or upto thousands of molecules innovated from the discovery group, maybe one to two of them get to be progressed into Phase III clinical development.

As mentioned earlier, the entire drug development process (both pre-clinical and clinical) can be viewed as a “weeding out” process – for every drug candidate, at every step of development, a “Go/NoGo” decision is made as to whether this candidate can be progressed into the next step of further development. Typically, the decision is based on the scientific evidence about drug efficacy and drug safety of this candidate. Throughout the entire process, if a drug candidate has any problems in formulation, toxicity, PK/PD, efficacy, or safety, then this molecule is weeded out from further development. Only the best candidates with good potential to become a successful drug can be progressed into the late phase clinical development.

During the clinical development of a new drug, there are lots of milestones and decision points. Many of these key decisions could have long-term impact and involve a large amount of resources and investment. As indicated earlier, the Go/NoGo decision after a PoC study is one of such examples. A NoGo decision means that all of the investments and efforts in developing this candidate up to early Phase II have not paid off. The high hopes of so many scientists and other project team members will not be realized. On the other hand, a “Go” decision implies a significant amount of additional investments and resources, as well as more man power will be committed in further development of this candidate. Important activities such as purchasing large amount of raw materials for preparing Phase III clinical trial supplies, as well as long-term toxicology studies, will be initiated.

However, such a Go/NoGo decision is not an easy decision. Usually it takes time to make a clear decision after the PoC study final data read out. When this is the case, the time between PoC results and a final decision is considered as a period of “inconclusiveness.” In addition to such major decisions, other Phase II considerations can be difficult as well. For example, how to choose doses for the upcoming clinical trial design? Which observed doses can be thought of as “efficacious”? What type of dose-response relationship

can be considered in moving forward to the next step of development? In Ting [5] the inconclusiveness regarding PoC decision will be discussed in further detail.

One popular confusion of evaluating a clinical project team's success is the use of the candidate product's success. In other words, some people tend to say a team is successful because the product is successful. Accordingly, if a product fails, the impression is that the team failed. This is, in fact, a misunderstanding. In clinical development of new drugs, the criterion to evaluate the success of a team could have nothing to do with the success of the product. An efficient project team should allow the good product to reach patients as soon as possible and should stop developing a potentially unsuccessful drug as early as possible.

On this basis, the criterion to evaluate the success of a team should be "re-work". In certain cases after a PoC trial is un-blinded, the team cannot make a decisive "Go/NoGo" call. When this happens, the team may have to design a second PoC – change doses, change patient inclusion/exclusion criteria, change sample size, . . . or other changes – in order to answer this "Go/NoGo" question. This is an example of re-work. In clinical development, re-work does not only cause delay in decision making but can also be very costly. The patent life of a product is 20 years. The time to develop a product can run at least 7 or 8 years and, in many cases, this time could be much longer. A re-work of one Phase II clinical trial can easily erode the patent life by at least one year. The amount of potential revenue of a successful product could run up to billions of dollars or more each year. Hence the delay caused by re-work can be very expensive. Major re-work because of unable to make a "Go/NoGo" decision could be considered as a failure of the project team, regardless on whether the product being developed will eventually be successful or not.

In designing clinical trials and sample size calculations, statisticians tend to pay attention to risks of developing a placebo (Type I error), risks of giving up a potentially good drug (Type II error), without paying attention to risks of inconclusiveness. From a new drug development point of view, risks about inconclusiveness should be considered before designing any clinical trials. During Phase II, when sample size is limited, it is very important for study design team members to notify upper management that when n is small, it is not only increasing the risk of making a wrong decision (Type I and Type II errors), but also the risk of inconclusiveness can be increases. However, even with large sample size, if the team is not clear about the confirmatory or exploratory questions, there can still be a rather high risk of inconclusiveness.

7. MODELS AND ASSUMPTIONS

Dose-response models are frequently used in the design and analysis of dose-finding clinical trials. In real world clinical development process, the introduction of dose-response

model actually increases the risks of new drug development. These increased risks can be thought of in a few various aspects. The highest risk is about model assumptions. The danger is that for those who proposed models, they themselves believed in their models. In the real world experiences of dose finding for new drugs, disasters after disasters happened because the model they believed is wrong. As pointed out by the renowned statistician George Box – "All models are wrong, some are useful". In dose finding clinical trials, all models are useless and misleading, in addition to being wrong.

In Phase II/III clinical trials for drugs treating chronic diseases, most of clinic visits are outpatient visits; i.e., a patient went to the physician's office for evaluation, and after medical services, the patient left the clinic. Under this setting, study medications need to be pre-formulated (likely to be tablets or capsules) with the protocol specified dose or doses. For example, if a dose ranging trial is designed with placebo, 25 mg, 50 mg and 100 mg – there are four treatment groups. Each patient is randomized to a given dosing group and stays with that same dose for the entire study period. The patient is instructed to take the randomized medication according to the time and frequency (once a day, twice a day, or other dosing frequency) specified in the clinical protocol. Test drugs of these dose strengths (0, 25, 50, or 100) have to be pre-formulated and packaged into blinded dispensable kits so that patients can carry these kits home, and take them according to the schedule.

For the following discussion, the focus is on the primary efficacy endpoint. At end of the study, after blind is broken, there are four groups of data – primary efficacy measures observed from patients receiving placebo, 25 mg, 50 mg, or 100 mg of study medication treatment. A summary statistics (mean, median, proportion, hazard, . . .) can be calculated for each one of these four groups. Note that these four summary data points are all that have observed from these patients. Once a statistician sees these four points, she/he has a tendency to "model" these points into a smooth, continuous curve. This tendency leads to one of the highest risks in designing and analyzing dose ranging clinical trials. The actual problem is not only the model itself, it is also about the unknowable assumptions behind such a model. The danger is that the statistician that came up with this model believes in this model without checking assumptions behind the model.

Note these four point estimates are discretely observed summaries, they are not from a continuous underlying model. One important realization is that there is no way to formulate every possible dose strength. When the drug is approved and marketed, patients are likely taking one or very few approved dose strengths. The thinking of a continuous dose-response model is not realistic, not necessary, and can be misleading. In dose finding clinical trials, the only available dose strengths are those pre-formulated discrete doses. A continuous dose-response curve (or model) never

existed in reality. Suppose the pre-formulated dose strength is 5 mg. Then the study team can consider testing 5, 10, 15, or any multiples of 5 mg dose strength. But there is no way to test 3 mg or 7 mg doses.

When making recommendations for doses of Phase III design, descriptive statistics are more reliable and trustworthy than doses recommended by models. When making dose recommendations for Phase III designs, statisticians can either simply use the descriptive statistics, or use confusing dose-response models. In most cases, the recommended Phase III dose(s) are the same by both methods. If this is the case, the effort to go through modeling is an entire waste of time, labor, effort, and talent. However, in some other cases, the recommended Phase III doses are different. In this situation, it is almost certain that the model is wrong. In practice, most of non-statistical team members only see these discrete data points, and they are not used to imagine a continuous dose-response curve. When working with non-statistical team members, it is important that the trial statistician does not confuse them. The thinking of models and assumptions from statistical training is actually a disadvantage for statisticians to communicate with non-statisticians.

Another risk of using dose-response models is the difficulties in communications with non-statisticians. As mentioned above, most physicians are not used to think of abstract objects. At study design stage, the statistician needs to convince physicians and other team members that this model is reasonable. By the time the study completes, clinical data do not necessarily fit the model well. How can the statistician earn their trust? On the other hand, if the statistician gives up the belief of any model in designing the dose ranging trial, it actually helps to build the trust from non-statisticians. Only after statisticians forget about model, they then can learn to respect data – dose recommendation for Phase III trials should be data driven, not model driven.

A different huge risk associate with dose-response models is the concept of “optimal” dose. In Phase II dose ranging clinical trials, if the statistician thinks about an optimal dose, the designed dose range could be too narrow. This is because the belief that for a given study drug, there should be a single dose that is optimal. With such a belief, the dose range tends to be narrow at the study design stage. One common situation happens in failure of Phase II dose ranging trials is the designed dose range being too narrow. Therefore, many tragedies in drug development originates from statisticians blindly believe there is an underlying dose-response model.

Suppose after data analysis, a model suggests the “optimal dose” is 17.6 mg. The existing formulation is tablets with 10 mg strength. For Phase III study design, the practical recommendation would be either 10 mg, or 20 mg (two tablets). It is important to note that an “optimal dose” learned from a dose ranging trial only reflects the optimality from a population point of view. However, in medical practice, the new drug is developed to treat individual patient.

It is well known that variabilities among human beings are huge. This “optimal dose” for a patient population can be far from the “optimal dose” for any given patient. Therefore, at the prescription time, the physician either prescribes 10 mg, or 20 mg for the patient to use. The idea of “optimal dose” is useless, and can be confusing. In drug development, it causes more un-necessary difficulties.

This is why models, and the thinking of optimality can be very harmful in the design and analysis of dose ranging clinical trials. In practice, we see failures in drug development can cost millions to billions of dollars in pharmaceutical industry, and the root cause could be these models. In most of cases when there is a major disaster in drug development, after the “lessons learned” discussion, the key mistake is not because “we did not know”, rather, it is because “we thought we knew”. That means, one of the root causes of major disasters is “assumption”. In the business of new drug development, most of drug candidates failed. Hence people in this business learned to be optimistic, to keep the hope up. But in reality, this optimism leads people to formulate unrealistic assumptions. When applied statisticians check their assumptions carefully in designing dose ranging trials, they will learn that dose-response models can be very harmful.

If statisticians designing dose ranging trials can be more humble in making assumptions, can develop the respect to data, and give up their belief of dose-response models, the entire pharmaceutical industry may save huge amount of wasted investments.

8. DESIGN CONSIDERATIONS

In order to simplify considerations in designing a clinical trial, the first point is to understand that the most important objective of almost every clinical trial should be to make a “Go/NoGo” decision. It is about whether the drug manufacturer, or the sponsor, should make further investments to continue develop this drug candidate. In practice, complications usually happen when this objective is not well understood, and not well thought through. This manuscript attempts to clarify why and how most of trial design questions should be linked to this basic question – should the sponsor continue to develop this drug candidate?

In the experiment design courses offered by most statistics or biostatistics departments, students learned that by focusing on a single, well-defined question for a given design, the results are more likely to be informative and meaningful. It is known as “one design answers one question” [7]. This principle can be studied from many prominent statisticians including RA Fisher [3] and J Neyman. The similar principle echoes by [1] that “the design of every clinical trial starts with a primary clinical research question”. In every clinical trial for new drug development, this primary question should be a “Go/NoGo” decision.

In fact, the most important deliverable of every clinical trial should be a “Go/NoGo” decision regarding whether

to make more investment to further develop this drug candidate. Drug developers need to distinguish two types of failures – 1. The drug candidate can fail because of lack of efficacy, or toxicity; 2. The study can fail because a clear “Go/NoGo” decision cannot be made. These decisions are especially important in Phase II clinical development. In practice, there are many risks that lead to the failure of a clinical trial. Among all risks, the worst one is unrealistic assumptions. This manuscript only lists some of these common risks. It is not easy to cover all possible risks. Here is a brief list of high risks team members frequently have to face in designing clinical trials – human behavior, small sample size, members understanding of question, communications, inconclusiveness, models and assumptions. In practice, many more potential risks can also be associated with clinical trial designs.

Given so many challenges and risks of designing a clinical trial, sponsors are still willing to invest in tens or hundreds of millions of dollars for the clinical development of new drugs. The important question is how the study team should think when designing a trial? Of course the trial needs to address the primary objective – even with all the risks that could potentially lead to an inconclusive outcome. Therefore, the key principle of any clinical trial design should be simple and robust. That is, with human behavior, limited sample size, problems in communications and assumptions, after clinical data read out, the team can still make a clear “Go/NoGo” decision. If the design is simple and robust, the risk of inconclusiveness can be reduced.

Again, the greatest risk among all possible risks is about unrealistic assumptions. The mathematical and statistical training of statisticians enable them to think abstractly. However, this abstract thinking process requires assumptions. It is very easy for a statistician to imagine a dose-response model without critically evaluate the assumption behind the model. Because “all models are wrong”, it is very risky to employ models in dose ranging trials. After all, the design and analysis of dose ranging trials should be a non-parametric practice. Any unjustified assumption can only increase risks of making wrong decisions, or risks of inconclusiveness. As a clinical statistician, we should be humble and not making unnecessary assumptions. More importantly, we need to suppress our urge of using dose-response models. After all, it is the observed data from clinical trials to drive our decisions for the next step – it is NOT the model.

Team members, especially the trial statistician, need to have a clear understanding of the trial objective. The study objective can be expressed in a single question with a well-defined primary efficacy endpoint, or in a sequence of questions with multiple endpoints, multiple doses, or subpopulations. No matter how complicate the real situation is, the most important question should always be about a “Go/NoGo” decision. At this point, the trial statistician is expected to organize the clinical questions into a logic

sequence with clear priorities. Next to write out a statistical hypothesis corresponds to each clinical question or endpoint. Then assign alpha to this sequence of questions. It is critical that all these questions have to be logically consistent, and without any ambiguity. During this process, people can easily be distracted by unimportant details. Hence it is always critical for the statistician to bring herself/himself back to the fundamental question – “Go/NoGo” decision. Should the sponsor invest more resources to develop this drug candidate after results read out from this study under design?

Confusion usually happens when multiple questions, endpoints, doses, or subpopulations are associated with this trial. This is the time to prioritize these questions – what is the key “NoGo” criterion? If the clinical results fail to reject this null hypothesis, then it is a clear “NoGo” decision. Consequently the question or endpoint associated with this statistical hypothesis becomes the primary question or endpoint. After this hypothesis is rejected, what would be the next question associated with the “NoGo” criterion? Order these questions step by step, then the priority of design questions become easy, clear, and explicit. At every step, critically examine the assumptions behind each hypothesis – can they be verified? are they realistic? how many of them? did we think of these assumptions when we write this hypothesis? what have we missed? At every step, keep in mind that most of disasters happen not because “we did not know”, but because “we though we knew”. This is an easy trap everyone can fall in.

9. CONCLUSION

The guiding principles of designing any clinical trial should be SIMPLE and ROBUST. A successful pharmaceutical statistician should “ask more, think more, and do less” when designing a study. Many of the fancy models and complex designs can actually do more harm to a trial design, than they can help. The main issue is that statisticians are trained to focus on risks of Type I error and Type II error, without considering the risk of inconclusiveness. After learning the importance of controlling risks of inconclusiveness, pharmaceutical statisticians will see the key design principles should be simplicity and robustness.

From a statistical methodology point of view, in the design of a PoC combined with dose ranging Phase II clinical trial, the general recommended method would be the OLCT approach [8]. OLCT is ordinal linear contrast test. For a combined PoC and dose ranging trial, multiple doses and placebo are included in a first Phase II trial to achieve both “Go/NoGo” decision and dose recommendation objectives. OLCT is a single degree of freedom t-test, there is no need for alpha adjustment. Using OLCT which combines all test doses and placebo to make a “Go/NoGo” PoC decision avoided multiple comparison adjustments. It simplified the multiple dose groups into a single decision framework.

OLCT can easily be understood by non-statisticians. After PoC, if the decision is “Go”, then use point estimates of each observed dose to make further dose selection recommendations. There are only two assumptions required to use OLCT – assume MTD guessed correctly, and monotonic dose response. The only caveat of applying OLCT is that in the central nervous system (CNS) or anti-psychotic drug development, monotonic dose response cannot be assumed, and OLCT may not be appropriate. Other than CNS, OLCT is widely applicable to most of therapeutic areas.

In fact, the best statistics for any given clinical trial should be that the statistical thinking covers the entire study – from study design, study conduct, to study completion and report writing. But other than the statistician, no team member feels any “statistical burden”. This means that team members do not have to consider statisticians as “necessary evil”; they do not have to do much in order to “satisfy the statistical needs”. In case report forms, only required data are collected, no more, and no less. In statistical analysis plan, only minimal number of tables are proposed, lean and mean, no fat. A good statistician design simple and robust clinical trials.

REFERENCES

- [1] EVANS, S. Fundamentals of clinical trial design. *Journal of experimental stroke and translational medicine* **3**(1) 19–27 (2010).
- [2] EVANS, S. and TING, N. *Fundamental Concepts for New Clinical Trialists*. Chapman Hall/CRC (2015).
- [3] FISHER, R. *The Design of Experiments*. Hafner Press, A Division of Macmillan Publishing Co. Inc., New York, NY, 10022 (1971).
- [4] PINHEIRO, J. C., BRETZ, F. and BRANSON, M. Analysis of Dose-Response Studies – Modeling Approaches, Chapter 10 of book *Dose Finding in Drug Development*. Springer (2006).
- [5] TING, N. Phase 2 Clinical Development in Treating Chronic Diseases. *Drug Information Journal* **45**(4) 431–442 (2011).
- [6] TING, N., CHEN, D., HO, S. and CAPPPELLERI, J. *Phase II Clinical Development of New Drugs*. Springer (2017).
- [7] YE, J., TIAN, H., GUO, X. T. and TING, N. Clinical Trial Design – What is the Critical Question for Decision-Making? *Statistics in Biosciences* (2023).
- [8] ZHANG, Y., DENG, Q., WANG, S. and TING, N. A Simple and Efficient Statistical Approach for Designing an Early Phase II Clinical Trial: Ordinal Linear Contrast Test. In *New Advances in Statistics and Data Science* 179–196. Springer (2017). [MR3586117](#)

Naitee Ting. StatsVita, LLC.

E-mail address: tingnaitee@gmail.com