

Clinical Trials and Biostatistics

Young Jack Lee, Ph.D.

LSK Global PS

jacklee@lskglobal.com

Abstract

Clinical trials are governed by two principles. One is the operating principle and the other is the statistical principle. The operating principle consists of two principles, protection of rights and wellbeing of participants and transparency. The statistics principle also consists of two principles, minimizing bias and maximizing precision. I will focus on statistical principles in this presentation. At the end, I will give my advice to statisticians engaged in clinical trials.

I. Introduction

Treatment decision requires medical evidence that the chosen treatment is beneficial to the patient without harming the patient. Medical evidence must show that the treatment is safe and effective. However, not every evidence of medical benefits is the same in its strength. Quality of evidence may be ranked, and Green and Byar (1984) proposed the following hierarchy:

1. Anecdotal case reports;
2. Case series without controls;
3. Series with literature controls;
4. Analyses using computer databases;
5. "Case-control" observational studies;
6. Series based on historical control groups;
7. Single randomized controlled clinical trials; and
8. Confirmed randomized controlled clinical trials.

Green and Byar consider, and it is generally agreed, the evidence from randomized controlled clinical trials is the strongest. Anecdotal case reports and case series without controls do not have control groups. They are two weakest. Series with literature controls, analyses using computer databases, case-control observational studies, and series based on historical control groups, although medical evidence getting stronger in the order presented, all have weakness in control groups.

Except for anecdotal case reports and case series without control, all the study designs have control groups. This implies that the efficacy of the test treatment is

meaningful only when evaluated in relation to control treatment(s).

The strength of evidence, however, is determined only by the quality of control groups. The highest quality medical evidence is produced by clinical trials with randomized concurrent control groups.

Quality of clinical trials with randomized concurrent control groups can be ranked further by whether the trial is blinded or not. Double blind randomized controlled clinical trials are ranked above and considered the gold standard.

Uncontrolled or poorly controlled clinical trials cannot provide convincing evidence about the efficacy of the test treatment even if the trial produces a strong positive result.

Randomized controlled clinical trials have two critical elements, concurrent control group and randomization. Randomization is to assure comparability between control treatment and test treatment. and existence of control group implies that the efficacy of the test treatment is relative. Some evidence, such as rabies shots, can be absolute. Such a dramatic treatment efficacy cannot be expected any longer. Furthermore most treatable diseases have effective treatments. Therefore, most new drugs' efficacy has to be relative to existing treatments. It is a must to have the placebo control group for trials of chronic diseases without effective treatments.

Not long ago, a domestic phase III trial result was presented to me for my opinion. The trial was an uncontrolled one. The sponsor was considering to license in the product for marketing. I refused to endorse the trial result because of poor design. Eventually the sponsor dropped the idea of licensing the drug. It is very rare to encounter uncontrolled phase III drug clinical trials these days except for device trials.

Clinical trials may be roughly categorized into exploratory and confirmatory trials where an exploratory trial generates information necessary for confirmatory trials which must produce definitive conclusion about the treatment efficacy.

Testing a priori selected hypotheses is the key component of the confirmatory trial, but exploratory trials may not even have hypotheses. Planning, design, execution, and interpretation of exploratory trials are flexible: even data driven hypothesis testings are tolerated. Phase I and phase II trials are considered exploratory, and phase III trials are confirmatory in nature.

We will not discuss exploratory trials in this presentation but only topics related to

randomized controlled confirmatory trials, also known as phase III trials.

II. Definition of clinical trials

Clinical trials are defined variably with slightly varying focuses. Here are some sample definitions. Friedman, Furberg, and DeMets (1998) define the clinical trial as "a prospective study comparing the effect and value of intervention(s) against a control in human beings." Green, Beneditti, and Crowley (2003) state "... controlled trials are by far the best method available for addressing difficult and controversial questions in a way that minimizes distrust of the results." Clinical trials are scientific experiments involving human subject, according to Lee (1999). ICH-Harmonized Tripartite Guideline E6(R1) (1996) characterizes the clinical trial as "[a]ny investigation in human subjects intended to discover or verify the clinical, pharmacological, and/or pharmacodynamic effects of an investigational product(s), and/or any adverse reactions to an investigational product(s), and/or to study absorption, distribution, metabolism, and excretion of an investigational product(s) with the object of ascertaining its safety and/or efficacy. " ICH-Harmonized Tripartite Guideline E6(R1) (1996) further clarifies, "The clinical trial and clinical study are synonymous."

Friedman, Furberg, and DeMets' (1998) characterization of the clinical trial is the phase III trial. Green, Beneditti, and Crowley (2003) stress that controlled trials, namely phase III trials, are the only method to produce convincing results. Because statistics is the discipline specializing in planning, designing, and analyzing an experiment, Lee (1999) considers statistics is the basis for clinical trials. The above three definitions are focused on confirmatory clinical trials. ICH-Harmonized Tripartite Guideline E6(R1) (1996) defines the clinical trial more broadly.

Key words about clinical trials are: difficult and controversial questions; prospective; comparison; scientific experiment; interventions versus controls; safety and efficacy; and in human subjects.

III. Why do we need clinical trials?

When we undertake a clinical trial of a new investigational product (IP or drug or treatment for simplicity), there has to be scientific evidence that the drug is safe and effective. Such evidence is generally produced through laboratory experiments and preclinical studies. In the case of herbal medicine, such evidence is not scientifically organized but rather traditional and empirical. Regardless of how evidence of safety and effectiveness of a drug is established, the evidence is not admitted as medical evidence until the drug is tested via rigorous confirmatory clinical trials. Why?

Knowledge about the disease and its courses is tenuous, and outcome of its treatment is not predictable. Often we do not know if the treatment will help the patient. For example, the use of beta blocker, now a popular therapy for heart failure, once was considered to endanger heart failure patient's life. Medical knowledge is constantly evolving and upending.

Disease is constantly evolving, and treatment must also evolve. Some diseases untreatable in the past are now treatable while some disease treatable in the past are becoming difficult to treat. Leukemia belongs to the former category, and tuberculosis caused by super bacteria belongs to the latter category. The clinical trial is the only discipline to assess objectively and scientifically the value of a treatment in a given setting at a given moment.

Treatment decision increasingly relies on the probability of treatment success which is estimated from clinical trial data. Well-designed and executed randomized clinical trials will provide a reliable estimate of success probability.

IV. Goals of clinical trials

Clinical trials are categorized into exploratory and confirmatory clinical trials. Phase I and phase II trials are considered as exploratory trials, meaning phase I and II trials are not considered to produce definitive medical evidence but to provide steps to confirmatory trials or phase III trials.

A confirmatory clinical trial is to generate objective and scientific data from human subjects receiving the treatment of interest, and to produce medical evidence about the treatment. EBM (Evidence Based Medicine) is what modern medicine pursues, and evidence from clinical trials is the most important and valued scientific foundation.

The key phrase is "objective and scientific data." Objectivity is consequence of reproducibility of clinical trial results and transparency of clinical trial process. This goal is achieved by rigorously following principles of clinical trials.

V. Principles of clinical trials

Clinical trials are governed by statistical principles and operating principles. Minimizing bias and maximizing precision are two primary statistical principles. The principle of protecting participants' rights and the principle of transparency are two operating principles.

There are many examples of rights abuse of clinical trial participants in the early, but not so distant, history of clinical trials. Not to repeat such abuses, protecting participants' rights and well-being is now considered the most important principle of clinical trials. The other operating principle, the principle of transparency, is to ensure reproducibility of clinical trial results.

ICH Harmonized Tripartite Guideline E6(R1) (1996) is about the two operating principles of clinical trials. Essential documents should show how these principles have been observed during the trial. Breach of these principles is considered as a serious violation that can result in nullifying the entire clinical trial and legal actions. We will not address this topic in this presentation any more, because the main subject in this presentation is statistical principles.

ICH Harmonized Tripartite Guideline E9 (1998), guideline of statistical principles of clinical trials, is the primary statistical document for clinical trials, and it is all about bias and precision.

VI. Bias and its sources

Bias in clinical trials refers to non-comparability of groups with regard to patient characteristics assigned to each group.

Let's assume that both treatment and control groups are assigned N subjects randomly. Of those, G_T treatment patients and G_C control patients have good prognosis, and the rest have poor prognosis. Prognostic factors may be known or unknown.

We now consider two possible scenarios:

Treatment is superior to control: If $G_T < G_C$, namely if the number of good prognosis patients in the treatment group is smaller than the number of good prognosis patients in the control group, the superiority of treatment may be confounded by the difference in the number of good prognosis patients. This is a bias.

	Number of good prognosis patients assigned	Number of poor prognosis patients assigned	Total number of patients
Treatment group	G_T	$N - G_T$	N
Control group	G_C	$N - G_C$	N

Treatment is not any more effective than control: If $G_T > G_C$, namely more good prognosis patients in the treatment group than in the control group, then a false positive result is likely. This is another bias.

There can be three different sources of bias including the one presented above:

Non-comparable control group: Bias will result if treatment group and control group are not comparable in the distribution of prognostic factors. If the control group is not concurrent, for example historical, or concurrent but non-randomized, then treatment group and control group are likely to be non-comparable. ICH Harmonized Tripartite Guideline E-10 (2000) is recommended for further reading.

Biased assessment of outcomes: Assessing outcomes of treatment is subjective in most clinical trials. This can cause conscious or subconscious bias in evaluating the outcome. For example, if the investigator is positively leaning toward the treatment, then positive bias may be introduced in evaluating the treatment outcome. On the other hand, if the investigator has a negative opinion of the treatment under study, then the treatment outcome may be evaluated with negative bias. Knowledge of what treatment the patient is receiving can also influence patient's own assessment of the treatment effect.

Biased data analysis: Even if treatment and control groups are made comparable by random treatment assignment, exclusion of participants from the data analysis can be different between treatment and control groups, making two groups non-comparable in the end. Suppose that the treatment is not effective with toxicity. When such a treatment is compared to placebo, poor-prognosis patients receiving the treatment are more likely to drop out than poor-prognosis patients assigned to placebo, because the treatment is toxic while placebo is not toxic. Such differential drop-out will make treatment and control group different at the analysis stage, and treatment group look more effective than control group, a definite bias. Statisticians tend to believe that statistical regression methods can remove most bias. If unknown factors are responsible for non-comparability, then statistical adjustment for bias will not be possible. Even if factors responsible for non-comparability are known, statistical adjustment will not remove all the bias, because the functional relationship between outcome measure and bias is likely to be unknown. Good statistical analysis methods can not make non-comparable treatment groups comparable.

Biases in clinical trials can be reduced only by randomization, blinding, and intent-to-treat analysis.

Randomization is the most effective method of creating comparable control groups by eliminating conscious and subconscious bias and by balancing the distribution of prognostic factors known or unknown. There are a number of different methods for randomization. Good Clinical Practice requires that the randomization should be reproducible and its programming code must be secure. It is a generally recommended practice that the statistician preparing the programming code is not the same as the study statistician analyzing the data.

Biased assessment of outcome can be reduced by blinding patients as well as investigators. In fact, data analysts may have opinion about the treatment and the study. Such opinions may bias data analysts too. Thus statisticians stay blinded until data are cleaned and locked. The statistical analysis plan is finalized before breaking the random code. Investigators cannot be blinded in some trials. In such cases, it is recommended that the treatment outcome is evaluated by blinded investigators separate from treating investigators.

Biased data analyses can be avoided only when all the randomized patients are included in the data analysis. Including poor-compliance patients in the analysis may dilute the treatment effect, but such dilution may be better than concluding ineffective treatment as effective or vice versa. Analyzing only compliant patients, namely per protocol analysis, should be avoided as much as possible except for exploratory data analysis purpose.

VII. Variability and its sources.

The outcome of the treatment varies from patient to patient. No two patients are identical in prognostic factors. Some prognostic factors are known, while other prognostic factors are unknown. Genetic factors are example of unknown prognostic factors. There are many such examples.

For example, until late 1970s oncologists were not aware that estrogen receptor status of breast cancer is a prognostic factor for effectiveness of chemotherapy. Recently Herceptin (trastuzumab) is found to be effective only in 20 to 30 percent of breast cancer patients whose cancer cells overexpresses a protein called HER-2.

In addition to prognostic factors, pure chance mechanism causes random variation in the outcome measurement.

There are two ways of handling known prognostics factors in clinical trials. One is to account for those factors in the design stage by incorporating them into randomization process. The other is to treat them as covariates in the analysis

stage. Regression method is the method of choice for analyzing effect of covariates on the outcome, but its validity and usefulness is constantly challenged when applied to clinical trial data. Taking important known prognostic factors into account in the randomization process, therefore, should be preferred.

Effect of unknown prognostic factors on the outcome is confounded with white noise, increasing the variance of outcome data. Randomization should balance the distribution of unknown prognostic factors between groups. But there is also a substantial risk of maldistribution of unknown prognostic factors. For example, suppose that a unknown prognostic factor is present in 20 percent of patients and that 300 patients are randomized between two groups. There is about 15 percent chance that one of the two groups has more than 36 patients with the prognostic factor while the other group has fewer than 24 patients. Such a maldistribution of the prognostic factor can have a serious negative impact on the trial result.

Thus it is very important to look for possible unknown prognostic factors when the trial result is somewhat different from expectation or when the observed variance is larger than expected.

In the planning and design stage, however, effect of unknown prognostic factors is not distinguished from the white noise.

VIII. Precision and the sample size

Precision in clinical trials means the size of risk for erroneous conclusions. The smaller the size of the risk the more precise the conclusion from the trial. Reducing the risk of erroneous conclusions, however, is directly related to the sample size. A clinical trial of a large sample size will produce a definitive conclusion, usually in the form of a small p-value if the treatment is effective, but will be costly. Therefore the sample size is very often a compromise between budget, time, and desired precision or intolerable size of risk for erroneous conclusions.

There are two types of erroneous conclusions from clinical trials. Concluding ineffective treatment as effective is called type I error, and concluding effective treatment as ineffective is called type II error. Such errors take place, even if clinical trials are designed and executed properly.

Probability of type I error, α , and probability of type II error, β , have different roles in clinical trials. A small α protects consumers from ineffective drugs while a small β makes sure consumers benefit from an effective drug. It is a common

practice to choose $0.01 \leq \alpha \leq \beta \leq 0.2$. We will come back to this point later in this section.

A sample size formula for two group comparison is $N = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{\delta^2}$ where N is the sample size per group. A common choice of α and β is 0.05 and 0.2, for which $z_\alpha = 1.96$ and $z_\beta = 0.842$. σ^2 is the variance of the individual observation, and δ is the clinically meaningful difference between two groups if two groups are different.

When planning a clinical trial, finding proper value of σ^2 and δ for the given clinical trial is always challenging. There is no royal way although a number of different methods are available. An interim evaluation of the selected sample size is now an accepted practice in clinical trials. This topic is getting a lot of attention lately.

In clinical trials, there are always subjects missing outcome measurement due to dropout from the trial and other causes. Compliance is always a problem. A simple minded adjustment of the sample size is $N = \frac{2(z_\alpha + z_\beta)^2 \sigma^2}{(1-d)(c\delta)^2}$ where d is the proportion of missing data and c is the average compliance rate. If the missing rate is 20%, the sample size has to be increased by 1.25 times to maintain the same power. But if the average compliance rate is 80 percent, the sample size has to be increased by 1.56 times. Compliance is a more serious problem than missingness of the outcome observation.

Often we need a rule of thumb sample size calculation. Occasionally statisticians are asked, for example during meetings, what the total sample should be for 20% response rate difference, for example. A rule of thumb upper bound of the total sample size for a two group study is $8/\delta^2$. α is a two-sided 0.05 and β is 0.2. For a 20 percent response rate difference, for example, an upper bound for the total sample size is 200, or 100 per group. For a 15 percent response rate difference, an upper bound is 356, or 178 per group. When more details are given, the final sample size can be smaller. Such a rule of thumb sample size should give a good guidance for planning purpose off hand. A one group study size, namely a non-concurrent control study, is 1/4 of the corresponding two group study size.

When designing a clinical trial, we choose $\alpha \leq \beta$ without much thought. Significance level, α , is a probability of false positivity. The probability of false positivity, however, changes with the underlying probability of the true positivity. The table below shows how this probability changes with changing probability (0.1, 0.3, 0.5, 0.7, 0.7) of the true positivity for $(\alpha, \beta) = (0.05, 0.2)$ and $(\alpha, \beta) = (0.2, 0.05)$.

If the underlying probability of the true positivity is only 10 percent, and if the trial produces a significant result at $\alpha=0.05$, then the probability of false positivity is 0.360, not 0.05. In oncology, there are more negative studies than positive studies. Thus, the actual significance level can be higher than the nominal significance level in oncology. In fact, if the probability of positivity is 0.5429, the probability of false positivity is 0.05 for a statistically significant result with $\alpha=0.05$. The significance level of $\alpha=0.05$ makes sense only when the probability that the test drug is more effective than the control is 0.54 or higher. I am not proposing a smaller α in oncology trials, but we have to be aware that the nominal significance level is smaller than the actual significance level in some disease areas including oncology.

If the probability of true positivity is high, then the probability of false positivity is smaller than the nominal significance level. For example, if $(\alpha, \beta)=(0.2, 0.05)$, and if the probability of true positivity ≥ 0.8 , then the probability that a positive result is false is ≤ 0.05 , namely the true significance level of ≤ 0.05 for a nominal significance level of 0.2.

Probability of false positivity when significant for the given α						
α	β	0.1	0.3	0.5	0.7	0.9
0.05	0.2	0.360	0.127	0.059	0.026	0.007
0.2	0.05	0.655	0.329	0.174	0.083	0.023

When selecting significance level and power for the sample size determination, one must consider the probability that the intervention under evaluation is true positive. For example, many new drugs being developed in Korea are a modified version of a proven effective drug in advanced countries. To my view, it is a waste of valuable patient and financial resources to require the precision of trials of such drugs to be at the level of a brand new drug development. Namely, Korea Food and Drug Administration has to consider seriously to allow $\beta \leq \alpha$.

The sample size is a guide to the size of a clinical trial, not a straitjacket that tightly binds the design of a clinical trial. It is my Korean experience that sponsors fret themselves about possibility of the per protocol sample size not reaching the designed size. Once I submitted a ballpark number for a phase II clinical trial to an Institute Review Board (IRB) of a Korean medical center. The IRB statistician demanded a precise hypothesis and justification. One cannot predict the precise per protocol sample size, and the phase II trial is an exploratory study in nature. Nothing can be exact in clinical trials, and demanding exact numbers in clinical trials is a simple sign for inexperience and lack of confidence.

It is a folly to seek an exact solution to an approximate problem according to a prominent statistician (Meier, 1975). A clinical trial is a scientific experiment involving human subjects. Nothing can be exact in clinical trials because humans are fallible. Furthermore, ethical requirements and human rights consideration have precedence over science. Trying to pinpoint an exact solution in clinical trials can be a very large folly. After all, we are not seeking a statistical answer to a medical question, but a medical answer to a medical question. Statistics is simply there as a helper albeit an important one.

IX. Conceptual relationship between clinical trials and hypothesis testing

Statisticians insist hypothesis test setting for confirmatory trials. In fact, even hypotheses are categorized into primary and secondary hypotheses. Statisticians concentrate only on the primary hypothesis specially in the trial design stage.

The strength of the hypothesis testing approach is that it cuts through many general observations and gets directly to a medical premise which we want to prove or disprove. This is an inductive approach. With all the general facts and observations on the table, statisticians force investigators to generate important medical questions by prioritizing and focusing on a few important and doable issues. Selected medical questions are further prioritized to generate the primary medical premise. The premise is that the treatment is more effective than the control under a given setting.

This premise becomes the primary hypothesis to be tested. We assume that the medical premise is not true or that the treatment is not any more effective than the control under the study setting. This is the null hypothesis. If the trial data show that this null hypothesis is not consistent with the data, then the premise is considered to be true; namely the treatment is considered to be more effective than the control under the study setting.

If we are to take a deductive approach to proving that a new treatment is beneficial, it can take a very tedious and arduous process, taking a step of progress at a time, like A to B, B to C, ..., Y to Z. Each progress can be a painstaking step.

The hypothesis generation step is a education process for statisticians as well as investigators. It is important to ask a lot of questions such as "What are study variables?" or "What are you trying to achieve from this study?" or "What is the budget and timeline?" etc. Investigators have many, often too many, facts and observations. Statisticians should be able to guide investigators to the primary

hypothesis.

Once the primary hypothesis is established, the study design is determined. The null hypothesis and alternative hypothesis are usually denoted by $H_0: p_C = p_T$ and $H_A: p_C \neq p_T$ where p_C is the success probability of the control and p_T is the success probability of the treatment. The sample size is determined for given α , β and δ where $|p_C - p_T| \geq \delta$ under H_A .

When the null hypothesis is rejected at the given significance and the observed success probability of the treatment is higher, some insist that $p_T \geq p_C + \delta$ is proven. It is wrong. That the null hypothesis is rejected simply means that $p_C \neq p_T$.

Once the null hypothesis is rejected, the estimated difference is claimed as the margin of superiority over the control.

The topic of superiority test and non-inferiority test will not be dealt here. There is one point I would like to make about a statement in ICH Harmonized Tripartite Guideline E9 (1998). At the end of *Section 3.3.2 Trials to Show Equivalence or Non-inferiority*, there is a paragraph about inappropriate interpretation of a non-significant superior test: *Concluding equivalence or non-inferiority based on observing a non-significant test result of the null hypothesis that there is no difference between the investigational product and the active comparator is inappropriate*. This statement is to discourage non-inferiority or equivalence conclusion from small size and/or inadequate studies which are likely to produce non-significant result unless the difference is dramatic. Non-significant results from appropriately designed studies with a proper sample size can be interpreted as equivalence or non-inferiority. *Post hoc* power analyses for a given δ may be necessary to claim equivalence or non-inferiority from superiority trial data.

X. Advices and lessons

I started my clinical trial career in June 1977. Dr. Richard Simon of the US National Cancer Institute was my boss when I started. When I started my biostatistician life, the field of clinical trials was very new, and there was very little guidance. Over the years, I have learned many lessons which I share with you.

1. Make sure you can clearly state your medical hypothesis to be tested. A broadly stated medical question causes unnecessary confusions and conflicts.
2. Often you have more than one hypothesis. You should be able to state what is the primary hypothesis. All other hypotheses are secondary and you must be prepared

- to abandon them because of different design requirements.
3. Your primary outcome variables should be validated before undertaking the study. Do not trust experts' claim that new untested outcome variables (or tested in a small sample) will work for your clinical trial. You do not want to find out whether untested new innovative outcome variables work in your clinical trial.
 4. Four reasons of non-significant medical trials:
 - a. Size of the trial.
 - b. Poorly selected outcome variables.
 - c. Poorly defined patient population
 - d. Truly non-significant hypothesis.
 5. Everything that can go wrong will go wrong:
 - a. Recruitment will always be slower than expected.
 - b. There are always ineligible patients randomized.
 - c. Outcome variables that you hope work do not work.
 - d. There are always missing variables and subjects.
 - e. Key support staff always leave at the critical moment.
 6. The state of art data management system may break down in the middle of the study. Insist on using the proven system that you know works.
 7. Responsibilities and authorities of participating investigators should be clearly delineated and established before the trial begins. Even so, there will be conflicts. Publication and authorship should be agreed upon before the trial begins. When personality conflicts flare up, the only working guidance is the agreement reached before the trial.
 8. Regardless of the size of the study, one should be prepared to account for every patient you recruit for the study as well as all the patients you do not include in your trial.
 9. In clinical trials, things can change momentarily. Do not put off anything you should do now. The next moment you will face new problems. You have to know your territory and establish all possible contingencies.
 10. Know this: clinical trials are scientific experiments involving human subjects, and the end product is a collection of numbers. Who is best prepared to make senses out of a heap of numbers? And the ultimate goal to answer is a *medical question* (emphasis added).

Some of the above lessons may not apply any more specially in registration drug trials, but may still apply in investigator initiated academic trials.

Although somewhat repetitive, I would like to stress the following statistical requirement for successful clinical trials:

- Medical question motivating the clinical trial: important enough and clearly defined;

- Primary response variable: medically validated and accepted and capable of answering the medical question
- Definition of target study population
- Primary statistical hypothesis: clear from the medical question
- Sample size: statistically justified
- Patient allocation: proper randomization
- Statistical data analyses: preplanned data analyses and maintenance of comparability of treatment groups

XI. Conclusions for statisticians

When the clinical trial is over, statisticians are responsible for data analysis. There are many statistical issues such as incomplete data, missing data, compliance problems, definition of analysis sets, subgroup analyses, multiplicity problems etc. These issues will be addressed later in the course. I want to point out the importance of simple analyses although generally clinical trial data are complex.

The CONSORT statement emphasizes simple unadjusted analyses: *Present sufficient simple (unadjusted) summary data on primary outcome measures and important side effects so that the reader can reproduce the results* (JAMA 12/28/94-Vol. 272, No.24). Furthermore, I may add, simple analyses are more convincing than complex analyses.

Mathematical modeling is an important guide for understanding the nature of data. But it cannot be the substitute for a simple and easy-to-understand data analysis. This reminds me of a dictum attributed to George Box: *all models are wrong, but some are useful*.

I often use the term "ball-park" in the spirit of the dictum attributed to John Tukey (Meier, 1975): *an approximate answer to the right problem is worth a good deal more than an exact solution to an approximate problem*. Results from a more elaborate data analysis and results from a simple minded analysis usually do coincide, falling sort of in the same ball-park. But, on occasion, they do not agree, and it is safe to say something is amiss.

In general, I analyze the data using both parametric and nonparametric methods, and report results from nonparametric methods, which tend to be conservative. If results are inconsistent, namely not falling in the same ball-park, then I try to find out why they produce different results. According to my experience so far, different methods tend to produce fairly consistent results. In one case, I advised the investigator not to publish the result because the result depended on the data analysis method. It was apparent that the data could not answer the question we were investigating.

Clinical trial data require imaginative and yet simple minded approaches. Applying simple methods is particularly important, because they generally require fewer assumptions about the data, and because investigators can readily understand the result. Simple analysis methods will make possible to repeat the same analysis for repeat trials, a general US FDA requirement for new drugs in most disease categories.

Lessons for biostatisticians

1. Listen and ask obvious – and stupid – questions such as "Why are you doing the trial?" It is rare that physicians can answer the question to your satisfaction.
2. Never argue.
 - You are always wrong.
 - All models are wrong, but some are useful.
 - There are no right answers and wrong answers. Some sound more reasonable than others.
 - What makes you think t-test is inferior to rank test. Both are approximate methods. Do both.
 - After all, statistics is an axiomless science.
3. Speak their language. Both you and they speak alien tongues no one can understand. When they do not understand you, they think you are not smart enough to handle their problems. You will find they can never learn your language but you can easily learn their language. Even those who claim they know statistics know very little when the surface is scratched.
4. Earn their trust and confidence. It is not your mathematical ability or degree or intelligence they rely on you. It is your integrity and their trust. Say "No" firmly when you have to say "No", but not too often. Only self-confident people can say "No".
5. KISS: Keep It Simple, Stupid. Use the simplest possible method and speak with authority. (This does not mean not to use advanced methods. Use both, but report the result from the simple method while keeping the result from advanced methods in the pocket just for the case. When results from a simple method and an advanced method do not agree, do not assume that the result from the simple method is wrong.)
 - Stubbornness is not authority. If you have to talk about statistical assumptions and formulas to convince them that your analysis is correct, then you are failing.
 - Simple solutions solve complex problems.
 - Complex solutions confuse everybody.
 - There is only one take-home message. It always is a medical one.
6. Do not believe what they claim they know. You have to be always be prepared for situations where nothing they say will work works.
7. Murphy's law: Anything that can happen will happen. Be ready. Even be prepared to

be fired.

8. You are free to express your opinion in everything. Do not forget, however, you are not paid for your medical knowledge but for your expertise in and knowledge of statistics.

In the perfect world, the clinical trial design is so good, and its implementation is so perfect that clinicians do not need statisticians to understand what has happened in the clinical trial. In the real world, the clinical trial is as perfect as human being is. Perfect clinical trials do not exist. Statisticians, however, should strive for perfect trials.

References

Friedman, L.M., Furberg, C.D. and DeMets, D.L. (1998). *Fundamentals of Clinical Trials*. Third Ed. New York: Springer-Verlag.

Green, S., Benedetti, J. and Crowley, J. (2003). *Clinical trials in oncology*. Second Ed. New York: Chapman & Hall/CRC.

Green, S.B. and Byar, D.P. (1984). Using observational data from registries to compare treatments: The fallacy of omnimetrics. *Stat. Med.* 3:361-370.

ICH Harmonized Tripartite Guideline E6(R1) (1996). Guideline for good clinical practice. Current step 4 version. *International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use* E6(R1).

ICH Harmonized Tripartite Guideline E9 (1998). Statistical Principles for Clinical Trials. Current step 4 version. *International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use* E9.

ICH Harmonized Tripartite Guideline E10 (2000). Choice of Control Group and Related Issues in Clinical Trials. Current step 4 version. *International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use* E10.

Lee, Y.J. (1999). Biostatistics and clinical trials: a view. *J. of Stat. Plan. and Inf.* 78: 349-367.

Meier, P. (1975). Statistics and medical experimentation. *Biometrics* 31: 511-529.

Piantadosi, S. (2005). *Clinical trials: A Methodologic Perspective*. Second Ed. New Jersey: Wiley-Interscience.