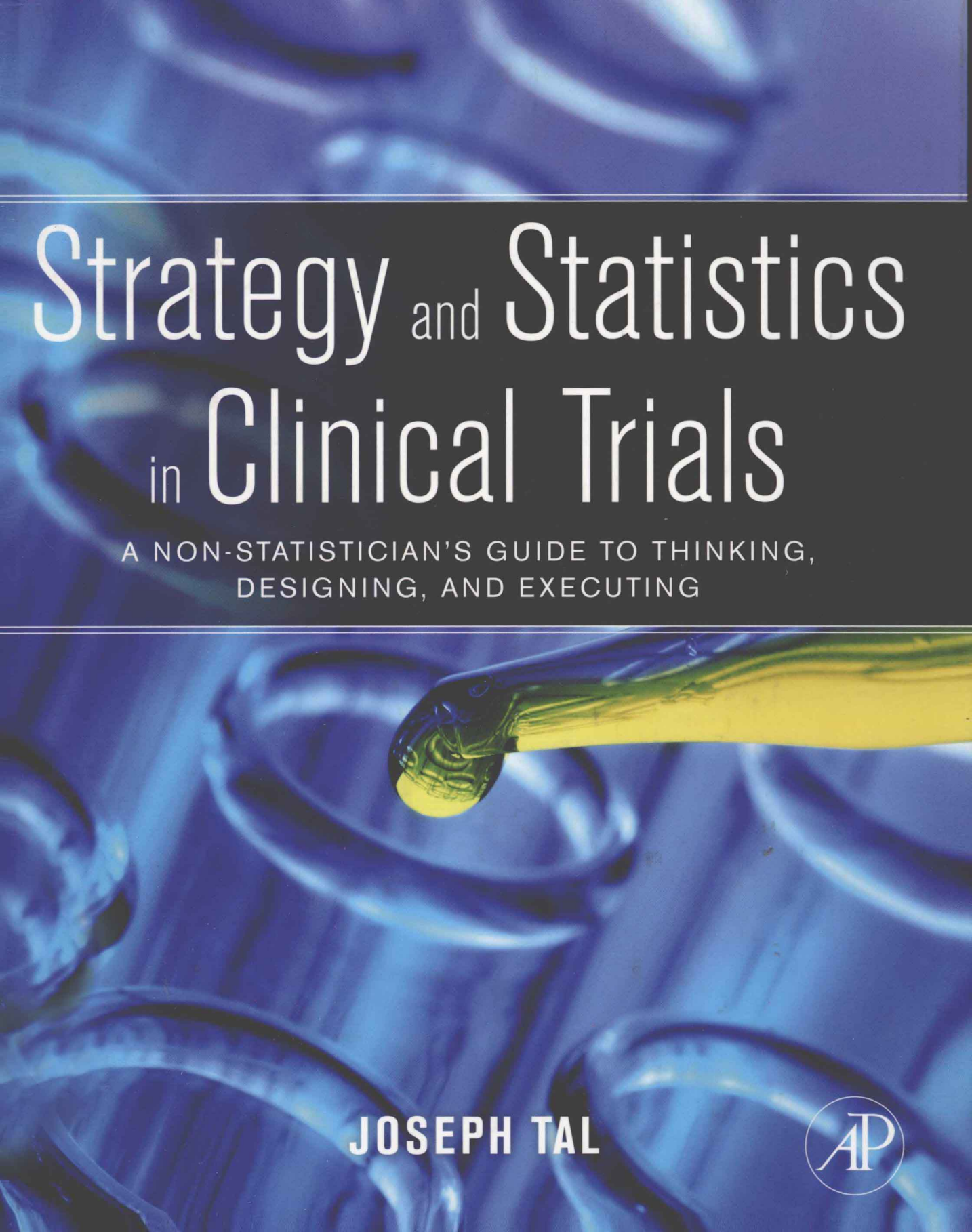
Strategy and Statistics in Clinical Trials
A NON-STATISTICIAN'S GUIDE TO THINKING, DESIGNING, AND EXECUTING
JOSEPH TAL
AP

# Strategy and Statistics in Clinical Trials

## A Non-Statistician's Guide to Thinking, Designing, and Executing

**Joseph Tal**



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK



**Notices**

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

**Library of Congress Cataloging-in-Publication Data**

Tal, Joseph.

Strategy and statistics in clinical trials : a non-statisticians guide to thinking, designing, and executing / Joseph Tal.

p. cm.

ISBN 978-0-12-386909-8

1. Clinical trials–Statistical methods. I. Title.

R853.C55T35 2011

615.5072'4–dc22 2011009725

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

For information on all Academic Press publications
visit our website at www.elsevierdirect.com

Printed in the United States of America
11 12 13 14 9 8 7 6 5 4 3 2 1

# Strategy and Statistics in Clinical Trials

# Introduction: Thinking, Designing, Executing

On those fleeting occasions when I regain my sense of wonder, I marvel at the complexity of things in general. During my own short lifetime we have put a man on the moon and observed ancient organisms on deep ocean floors. Computers and cell phones have become near necessities, and even inexpensive cars do not often break down (a minor miracle to my reminiscing mind). Being adaptable, we have become accustomed to the most elaborate of devices. And we take them for granted as one might have a kitchen knife some hundreds of years ago.

Life sciences have not lagged behind. We have mapped the basic structures of life and increasingly have come to understand the human body and the substances circulating in it. Using this knowledge, we have developed products that can alter and regulate both mind and body. To be sure, there is still much to be done, and hopefully there always will be. But we have achieved a great deal already.

Now and again I talk with people outside my field and find it both fascinating and frustrating. Fascinating because there are so many incredible developments out there, and frustrating because it is nary impossible for me to know enough to truly appreciate them. Our world has become so complex that it is difficult for any one individual to have little more than a single area of expertise. And even this can only be had with a great deal of effort. While I might wish it otherwise—and I do—this is the way it is. But more disconcerting is that specialization often impedes the work of people who *should be* working together: professionals from disparate fields who must team up to get things done.

My own area is statistics. Within it, I specialize in biostatistics. Specializing even further, I have gained some expertise in my discipline's methods in clinical trials. And when I return to Earth from my contemplative heights, I find there is often great disorder in these trials. At times it seems to me a wonder they even work at all. Clinical studies are planned and executed by dozens of people both within and without the organization. In great measure their success depends on

the good graces of harried physicians and volunteering subjects who are often very infirm. A single trial may be conducted across many centers, countries, and even continents, making its management that much more difficult. And studies often take years to complete, during which time anything can happen and generally does.

So like most complicated ventures, success depends on numerous processes and professionals with varied expertise. A partial list of specialties includes finance, clinical, regulatory, marketing, toxicology, biology, physics, materials engineering, software engineering, medical monitoring, analytic chemistry, and information technology. And yes, statistics. If we cannot coordinate effectively between these specialties, our trials will be suboptimal at best.

Paradoxically, our complex products demand both added specialization within fields and ever more dialogue between them. So while we ask our people to know more about less, we increasingly require that they communicate with others in the same predicament. And as specialization deepens, interactions between experts become more difficult.

Simply put, for a clinical trial to work people must talk to one another. And while it is impossible for any to fully understand all others, each must know enough for the dialogue to be useful. This then brings me to my book's objective.

Statistics—be they more complex or less—are involved in virtually every clinical trial. The discipline provides essential input in the planning stage on issues like trial design, choice of endpoints, and sample size. And it supplies the language for communicating outcomes using simple statistics like mean and median, and more sophisticated tests for inferring conclusions. This is my work, and I like it. And in medical research, statistics is important work. But just because it is, I do not expect others to take year-long courses in it. Indeed, among those who already have, many are perfectly happy to leave this year behind them.

For a large number of clinical trial professionals my discipline is a black box they are content to leave as is. But this must not be if we are to design, conduct, and report trials effectively. In the chapters that follow I aim to cast some light into this box.

This book explains clinical trial statistics in the simplest language I can manage. There is very little formal mathematics in it and almost no formulas. More importantly, I place the discipline in the wider context of clinical trials—the inevitable constraints of time and money, and limitations associated with clinical practice. I also relate trial design and analysis to its intended audience—to those needing the information it provides, such as regulators, scientists, physicians, corporate managers, investors, and others.

The book is a practical guide. It is based on years of applied experience, much of it my own. In it I present numerous examples from pharmaceuticals, devices, and other products. Crucially, I describe how statisticians must consider a trial's overall needs and reconcile to them. And I show how at times it must be the other way around. Be that as it may, a clinical trial cannot maximize any one discipline's preferences. But it *can* optimize—and it must.

For nonstatisticians this book provides strategies for productive dialogue with those who are; it describes the statistician's approach to clinical trials and the basic tools at the statistician's disposal. For statistical professionals this is a "how-to" guide for interacting with the many others working on clinical studies. In this book I describe—for the benefit of statisticians and nonstatisticians alike—the numerous considerations involved in clinical studies and their effect on a trial's statistics.

Competence and native intelligence will get you a long way in most every field. But little apart from experience can give you *experience*. Well, I present here what I believe to be the next best thing: other people's experience. I describe the central topics of clinical trial statistics with real-world examples, recounting clinical tales and their morals. I have to the best of my ability written a book to facilitate communication between the field I have chosen and those I have not.

# Table of Contents

CHAPTER 1

# Clinical Development and Statistics: The General View

## CONTENTS



## INTRODUCTION

When putting together a clinical trial, each discipline involved brings its own particular and peculiar view to the table. Some of these are complementary, while others combine less seamlessly. Still others will pull in different directions. It is your job to find the best way to profit from this kaleidoscope of views—to merge the varying approaches to produce a solid study.

*Clinical trials are a multidisciplinary effort.* Now this is a very general statement and is true of many professional endeavors. Besides, you have long known it to be true. So instead of keeping to generalities, let me begin with a hypothetical, though typical, example from pharmaceuticals. In it I will describe some of the central issues that arise when planning a clinical study and focus on the biostatistician's role in them.

I do not intend to be exhaustive here. It is virtually impossible to cover all relevant issues and, more to the point, it is impractical. My aim is to provide an idea of statistics' contribution to planning a clinical trial. This will be enough to keep me occupied for a while.

## ALLERTON'S PALSY

Allerton's Palsy (ALP), whether discovered by him or on him, is a **neurodegenerative** disorder primarily affecting physical function. In about 10% of cases there is internal organ involvement, and there is some question whether these cases constitute a distinct disorder. Be that as it may, ALP is primarily a disease in which the immune system misidentifies a part of the brain as foreign and attacks it—that is, an **autoimmune disease**. The brain cells (**neurons**) attacked by ALP degenerate over time. This usually leads to serious disability and in some cases to death. The **natural history** of the disease—the individual path it takes—is highly variable. Some with ALP show little change in physical functioning over decades, while others will deteriorate within a few years. Most commonly, the disease progresses slowly at first and accelerates over time.

Diagnosis of ALP is difficult so the number of reported cases is unreliable. The onset of the disease is usually between the ages of 30 and 45, though much later onsets have also been documented. ALP is more common in men than women, with an estimated male-to-female ratio of 3:1.

Your Company has developed CTC-11, a molecule designed to protect brain cells at risk from the disorder. As such, your compound aims to provide **neuroprotection** in ALP. Virtually all early studies with CTC-11 have been encouraging: **In vitro** testing demonstrated significant reductions in cell death, whereby animal neuron cultures treated and exposed to a toxic environment displayed a much smaller rate of cell death than those untreated. Other **preclinical** studies showed that the molecule delays onset of the disease in rats. In animals that had already developed ALP, CTC-11 slowed progression of the disease relative to those not treated. Preclinical testing also indicated that in the **animal models** tested, the molecule is safe. Finally, two early **Phase I clinical trials** in healthy volunteers showed that even high doses of CTC-11 produced only minor side effects, such as headaches and transient rashes.

So it seems you have in hand a potentially safe and effective drug for ALP—a chronic disease that currently has only marginally effective treatments and no cure. But it is still early and there are many issues to be considered before moving on to the first clinical study in humans with the disease. In the following sections I examine a sampling of these issues and articulate some typical questions that arise. And for each, I describe the statistical input required for providing answers.

### Science

In vitro testing is about as far from the clinic and its patients as you can get. Success at this stage is supportive of **efficacy** in humans but is often little more than that. The positive results observed in animals provide you with stronger support; after all, the molecule has now been tested in a living organism.

But here too the evidence is fragile. First, results in animals frequently do not replicate in humans or only partially replicate. Second, and more important, ALP is a human disease that does not naturally occur in any other species. Unfortunately, the animal models developed to study ALP have yet to be fully **validated**-scientists are unsure whether they mimic the human form of the disease well enough to serve as suitable models for it. So the best you can now say is that CTC-11 has effectively treated animals that have been caused to develop an ALP-like disorder. But has your drug actually treated *ALP*? Well, you would like to think so. Yet you will not know until the molecule is tested in humans with the disease.

Despite all the caveats, there is no denying that your results have been good so far. In fact, some of your outcomes in animals could even be termed impressive. And since you cannot do much better than you have, it is only natural that you want to know what CTC-11 can do for humans with ALP.

Thus your continuing efforts with CTC-11 are scientifically important and may end up helping some very ill people. If successful, they will likely contribute to your Company's coffers and, with a bit of goodwill, to your own modest ones as well. Finally, a successful effort will almost certainly get you that lunch with Aunt Augusta promised in a moment of weakness—the one contingent on your "finally doing something useful with your life."

Your Division Head has asked you to prepare a presentation for management, which is now ready to decide on the molecule's future. Management, she says, would like a summary of the results obtained thus far. She gives you about a week for it. You work late hours and a weekend to convert the large amount of information generated to date into a presentation that even management can understand. You contemplate incorporating something about the importance of the project to patients and a word or two about market potential as well. But you are an R&D manager and your boss has made it clear that the presentation should deal with scientific findings only. "Stick to the facts as you know them," she says, and you wonder which version of the Bible they will have you swear on.

The presentation is done. After more than three hours of endless chatter (only a small part of it your own), it is over. Slightly sweating still, you leave the room and close the door behind you. The air in the hallway feels surprisingly fresh compared to what passed for oxygen in the conference room. You take the stairway down and do not even notice the receptionist's benevolent smile as you wait for the automatic door at the entrance to open. You go through the door and out, and for a while your mind is empty.

Walking to your office you think back to those long hours in the lab—the many frustrations and few rewards. And you ponder the time it took to prepare the presentation with every sentence and chart in its proper place. Did anyone notice? You certainly hope so.

Clutching the laptop as if it were your only friend, you wonder if it all comes down to this: three exhausting hours and a PowerPoint file. Of course you know that a management presentation alone will not determine the fate of a development program. But that is how it feels.

Well, you have done your best and believe you have made a good scientific case for the molecule. Your Division Head backed you throughout the meeting, as did a couple of VPs. The CEO appeared positive, but he has this way of nodding his head that makes you think he would have preferred to shake it instead. You simply do not know. Regardless, other than one wearisome VP nearing retirement, no one was outright negative. And now they must decide whether the scientific evidence supports the substantial investment needed for testing CTC-11 in patients with ALP.

### *Statistical Input*

The preclinical studies were likely conducted with the statistician's help in data analysis and have his stamp of approval. With any luck he was also involved in designing these studies, which was almost certainly the case in Phase I.

But all this is behind you. The evidence from these early experiments is in its final form, and just about all that could be learned has been learned. It is now left for others to decide if R&D has presented sufficient evidence for CTC-11 to justify the cost of moving forward.

Still, the statistician might have helped you prepare your presentation by pointing out which numbers are best presented and how. He may also have suggested a graph or two and some color-coordinated charts. But beyond what he has already done in CTC-11's development, there is little of substance for him to do now. In short, the statistician's input into the scientific aspects of results *after* they have been obtained and analyzed is limited.

There is, however, an exception that typically occurs in **due-diligence**, where individuals consider investing in a particular R&D venture. These potential investors may seek a statistician's evaluation of the *quality* of the evidence obtained—the degree to which the results presented were produced by well-designed and correctly analyzed studies. Clearly, well-designed studies provide more reliable conclusions than sloppy ones, and correct statistical analyses are more apt to yield an accurate picture than incorrect ones.

## Trial Sample Size

After numerous management meetings and a discussion with the board, the Company has decided to go ahead with the molecule. Public Relations have come out with a press release emphasizing the favorable results obtained to date and stating that the "Company is excited about pursuing further

development of CTC-11, a promising drug for Allerton's Palsy." The drug, states the press release, "will enter a **Phase IIa clinical trial** in the first quarter next year," and there is some hint in the release of a pivotal **Phase III clinical trial** to come. But you are so euphoric with the current decision that the mention of Phase III makes little difference. Besides, you have learned enough about press releases to know they are not meant for you. When doing science, it is better you read Excel sheets than broadsheets.

So the Company will pursue further development of CTC-11. However, as always, you will be testing the molecule with limited resources. At this stage the budget is for a 6-month trial with about 150 subjects; half will receive CTC-11 (**Treatment group**) and half will receive **Standard-of-Care (Control group)**. Will this be enough?

### *Statistical Input*

Many compounds do not go directly from Phase I into full-fledged trials like the one proposed. A smaller pilot study is probably more common and, under the circumstances, perhaps more advisable. In fact, you have no idea where the number 150 came from and suspect it had more to do with budgets and stock prices than with the development program's needs. Regardless, this is what you have been given and it is substantial. But "substantial" does not necessarily mean "sufficient," the relationship between the two depending on the case at hand.

Numbers—as large or small as they may seem at first—cannot be evaluated without a context. A 9-year-old child selling lemonade in front of her family's garage might feel that taking in $30 on a single Sunday makes her the class tycoon. But offer her the same in a toy store and she might complain of underfunding (and, frighteningly, might use these very words).

Be that as it may, this is what you have and you must make the best of it. Still, you are not going to take the numbers proposed as set in stone, and one of your first questions is whether they will provide your development program with the information needed. Specifically, will this study produce enough data for making an informed decision on taking CTC-11 into the next level of testing?

The statistician's role here is central. He will likely begin with straightforward **power analyses**, which here relate to calculations determining the number of subjects needed for demonstrating the drug's efficacy.[1]

---

[1] This is assuming the drug is effective. If a drug is ineffective, no sample size will make it otherwise. Thus, the goal of power analysis is to compute a sample size that will provide statistically significant results given a product that is assumed (or known) to have a particular level of efficacy.

In a future chapter we will deal with power analysis in greater detail. For the moment let us point out that to do these analyses a statistician needs several pieces of information. The most important of these is an estimate of the drug's **effect size** relative to Control. For example, stating that CTC-11 is superior to Control by about 10% is saying that the drug's effect size is about 10%.[2]

The statistician will get these estimates from clinicians and others in the organization. But he should also review results obtained to date within the Company and read some scientific publications on the subject. To do this he will need assistance from life-scientists, without whose help he will have difficulty extracting the required information from medical publications.

This is but one example of professionals from different fields needing to interact in trial planning. In this book I will note many more. So while statisticians need not have deep knowledge of biology or chemistry or medicine, they should be sufficiently conversant in these disciplines to conduct intelligent discussions with those who are. And the same goes for life-scientists, who would do well to be conversant in statistics.

Once acquired, the statistician will incorporate this information into his power analyses. These will yield sample sizes that will be more useful than those proposed primarily on financial considerations. If management's proposal and the power analyses produce very different sample sizes, you will (alas) have another opportunity for multidisciplinary interaction.

## A DIFFICULTY WITHIN A PROBLEM

You have asked the statistician to compute the required sample size that will ensure your trial is a success: the number of subjects that will provide sufficient information for making future decisions on CTC-11. The statistician, in turn, has asked *you* for information; he has requested that you estimate the effect of the drug relative to Control. On the face of it, this is a silly request. After all, you are planning to conduct a trial precisely to discover this effect, so how can you be expected to know it *before* conducting your trial? To tell the truth, you cannot know it. But you *can* come up with an intelligent guess and have no choice but to do so. Indeed, estimating an effect size for the purpose of planning a trial of which the purpose is to estimate effect size arises often. We shall deal with it later, but for the moment let me assure you it is not as problematic as it sounds.

When determining sample size, the statistician will do well to talk with physicians and marketing personnel regarding the kind of CTC-11 efficacy needed for the drug to sell. Incorporating this information into power analyses will provide the Company with data on how valuable (or not) trials of varying sizes are likely to be from the standpoint of assessing market need.

---

[2] Quantification of effects—effect sizes—come in a variety of forms, percent being one of several.

The statistician should also expand his exploration to alternative study designs—not just the initially proposed six-month study of 150 subjects in two **arms**. Some of these designs will require fewer resources, while others will require more. He might, for example, examine a scenario where the larger trial is replaced with a smaller **pilot study** of 10 to 30 subjects. This sort of study could provide a more realistic estimate of the drug's effect in humans—an estimate that is now lacking. Once the pilot study is done, there will be more reliable information for planning the larger trial.

The larger the trial, the more informative the data obtained from it. But, as Goldilocks demonstrated years ago, strength does not necessarily reside in numbers; if a smaller trial can provide us with the required information, we should prefer it to a larger one. Conversely, if the larger study has little potential to provide the required data and an even larger trial is needed, you would do well to forgo the former and request more resources.

So a small pilot may be just what the statistician ordered. But this pilot will come at a price: A two-stage approach—a pilot and subsequent, larger trial—will slow down the development process. Moreover, given the fixed budget, any pilot will come at the expense of resources earmarked for the second stage. Here too there is more than one option. For example, you can design a standalone pilot and reassess development strategy after its completion. Alternatively, you can design the larger study with an early stopping point for **interim analysis**—an early check of the results. Once interim results are in, the information can be used to modify the remainder of the trial if needed.

These two approaches—one that specifies two studies and another that implies a single, two-stage study—can have very different implications for the Company. They differ in costs, logistics, time, flexibility, and numerous other parameters. The choice between them should be considered carefully.

For the moment let us simply state that the statistician's role is central when discussing trial sample size—the number of subjects that should be recruited for it. At the same time, it is very important for those requesting sample size estimates to actively involve statisticians in discussions dealing with a wider range of topics as well—for example, the drug's potential clinical effects and alternatives to the initially proposed design. And given that it takes at least two to trial, it is critical that the statistician be open-minded enough to step out of his equation-laden armor and become cognizant of these issues.

In sum, the fact that a relatively large sample size has been proposed for this early trial does not necessarily imply that it will provide the information needed. Together with your colleagues in R&D, logistics, statistics, and elsewhere, you should discuss all realistic alternatives: There can be two trials instead of one, one two-stage trial, as well as trials with more than two arms or less, a longer trial or a shorter one, and so on.

Now all this may seem a bit complicated, and it can be. At the same time you should keep in mind that because your budget is limited, the universe of possibilities is restricted as well; covering all, or nearly all, study design possibilities given fixed resources is definitely doable.

### *Trial Length*

Virtually all published studies investigating treatment for ALP evaluated effects of anti-inflammatory agents, both steroidal and not. Where positive results were obtained (albeit modest ones), they appeared between three weeks and four months into the trial. But CTC-11's hypothesized **mechanism of action** is neuroprotection; it is meant to shield brain cells from processes leading to degeneration and death. **Neuroprotective** activity is most useful in the long term, with its short-term effects likely to be more subtle than those of anti-inflammatory agents. As a result, your new drug may even prove *inferior* to existing treatments in alleviating short-term symptoms, such as pain and fatigue.

So you reason that your best bet for success may be demonstrating CTC-11's effect on Disease Progression rather than short-term alleviation of symptoms. Of course, in any study conducted, you will also collect data on symptom relief. It is just that you feel that longer-term measures will highlight CTC-11's benefit more than short-term ones.

By definition, **endpoints** associated with Disease Progression measure change over time. In chronic diseases these endpoints generally reflect the deterioration that occurs with varying rates, depending on the specific disease, individual patient, quality of treatment, and other factors. One measure of Disease Progression in ALP is the Mannheim Working Group Lower Limb Reflex Response Scale (MLRS). The measure consists of multiple items, most based on a physician's exam of foot and knee reflexes and several reported by the patient. The MLRS is relatively **reliable** and has shown at least moderate relationships with other important disease parameters.

Based on a review of scientific literature you conclude that meaningful declines in MLRS in ALP typically take at least nine months to appear; they are rarely observed in six months or less. This is especially true for patients in early stages of ALP, where progression is slower than at later stages. And your initial trial is intended for early-stage ALP patients.

Based on these data and CTC-11's mechanism of action you believe that a longer trial will increase your chances for success. Consequently, you feel that the MLRS—a measure sensitive to Disease Progression—should be the most important efficacy endpoint in your trial. It will be the study's **Primary Efficacy Endpoint**.

Now it might seem that I have moved away from sample size and study design and entered a discussion of endpoints in a clinical trial—in other words, that I am now dealing with parameters, such as MLRS, that will ultimately determine whether one drug is better than another. This is indeed a very important discussion, but I have *not* entered it despite clear indications to the contrary.

Deciding on a trial's endpoints, particularly the central ones, is critical and will be dealt with in time. But in this section I focus on trial length and will ignore the issue of endpoint selection to the best of my ability.

## A MORAL

In clinical trials, as in life in general, just about everything is connected to everything else. So it is that you wish to limit your discussions to trial length and find yourself slipping ever so naturally into sample size computation and endpoint selection. For many of us, it would be nice if both life and science were to proceed in an orderly fashion. There would, as the Greeks taught us, be a "beginning," "middle," and "end" to everything. But the Greeks were strong on mythology, and this concept of orderly progression is often only tenuously related to reality.

Issues and events generally do not advance in the uniform and orderly fashion we would like. And even scientists have questioned the established order. Thus, for example, where evolution was once thought to proceed slowly and at a relatively even pace, this no longer is the consensus. Oddly, artists have found themselves in a similar seafaring vessel. Thus, a good many artists in the last century created works of which the point was that neither time nor space is arranged in a particularly orderly fashion.

The division of larger concepts into smaller, more manageable ones is often artificial. But I will make an effort at order regardless. It is how we learn best and, at this stage, we are learning. At the same time, I suggest you keep in mind that the process of planning a clinical trial typically involves at least as much disorder as order. And the confusion ebbs and flows as we design and execute the trial while dealing with finances, logistics, people, and unexpected events that arise to produce ever more fascinating forms of disorder.

Now let us get back to the trial. Summarizing thus far, you are contemplating a study showing CTC-11's effect on Disease Progression, feeling this is where the drug's greatest advantage is. And you believe MLRS is the best measure for demonstrating it.

But MLRS can only pick up relatively large changes in Disease Progression, and these take more than the six months currently planned for your trial. So if you want to show meaningful Treatment-Control differences on this particular endpoint, you will need a longer trial, which will present its own problems. Longer trials require more resources than shorter ones and are more difficult to get right: The logistics are more complex, larger numbers of patients are lost along the way, and the number of unpleasant surprises popping up will be larger as well. Moreover, a 9- or 12-month trial will delay the Company's development program and, with it, your long-awaited lunch date with Aunt Augusta. On top of all of this there is a price to pay—literally. Given a fixed budget, extending the trial must come at the expense of something else. And this "something else" is probably important for the trial's success as well.