Practical Biostatistics in Translational Healthcare

Allen M. Khakshooy
Francesco Chiappelli

EXTRAS ONLINE

Springer

Practical Biostatistics in Translational Healthcare

Allen M. Khakshooy · Francesco Chiappelli

Practical Biostatistics in Translational Healthcare

Springer

Allen M. Khakshooy
Rappaport Faculty of Medicine
Technion-Israel Institute of Technology
Haifa, Israel

Francesco Chiappelli
UCLA School of Dentistry
Los Angeles, CA
USA

ISBN 978-3-662-57435-5 ISBN 978-3-662-57437-9 (eBook)
https://doi.org/10.1007/978-3-662-57437-9

Library of Congress Control Number: 2018943260

Printed on acid-free paper

This Springer imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

To master and apprentice,

“Credette Cimabue ne la pittura tener lo campo, e ora ha Giotto il grido, … ”

(Dante Alighieri, Divine Comedy, Purgatory XI 94,95)

Foreword

It is a great pleasure for me to write the Foreword of this interesting book on biostatistics and translational healthcare. Biostatistics is a rapidly evolving discipline. Large databases that collect massive information at individual level, coupled with high computing power, have increased enormously the potentiality of this discipline to a point that we can realistically speak of evidence-based patient-centered healthcare. However, the foundations of this discipline have not changed over time and lay on arguments that are inherently philosophical and logical. They also require a deep subject matter knowledge. All these issues are well addressed in the book, which constitutes a very good introduction to the principles of biostatistics.

The book describes in detail the research process underlying the construction of what can be considered the scientific knowledge. The research process is described as three-legged stool: study design, methodology, and data analysis. Each leg is investigated with profound competence. Carefully designed examples introduce the students to the core concepts and, at the same time, equip them with a critical view. I particularly appreciate the description of the sources of errors hindering any statistical conclusion. Random errors are by far the most studied and widely understood. Nowadays, also systematic errors, due for instance to self-selection, missing information, informative drop-out, and so on, have received much attention in the statistical literature, thanks to sensitivity analysis and related methods. Left behind, to my opinion, are what in Chap. 1 of the book is described as the most dangerous source of mistakes in research: the errors of judgment. Improper use of logic and information induces fallacies in the interpretation of evidence and the formulation of the conclusions. These types of errors are now even more common, since the inrush of the so-called big data, i.e., data that are collected from different sources with different purposes. Errors of judgment are subtle and subjective and cannot be easily transferred into a statistical model. For this reason, they are often overlooked.

Starting from the first leg, study design, the book describes the different studies that are relevant in biostatistics, ranging from diagnostic studies to research synthesis. It then focuses on the major distinction underlying prognostic research, namely, observational and experimental studies. Experimental studies with random assignment of subjects to the treatment are the ideal framework for research in biostatistics. However, ethical reasons together with practical obstacles make this study not feasible in many contexts. Moreover, complete adherence to the study protocol is rare, and elements of observational studies are therefore introduced even in a well-designed experimental study. It is therefore crucial to understand the potential sources of bias due to the absence of randomization. As the recent reprint, with comments, of the seminal paper by Cochran (1972) "Observational studies" witnesses, the issue is of crucial importance in all applied research areas, with biostatistics being a notable example.

The book then carries over by addressing the methodology. Emphasis is placed on the sampling procedures as well as on data acquisition through valid and reliable instruments. Coming to data analysis, descriptive statistics and inferential methods are presented, with an eye to the process that transfers research results into new methods for diagnosis, therapy, and prevention, that is, the object of translational healthcare. I particularly appreciate the emphasis on questions that a researcher must address to ascertain the internal and external validity of a study that constitutes the replicability of the findings and therefore their accreditation in the scientific community.

The final part of the book is dedicated to the consolidation of statistical knowledge and its capacity to "make a difference" in the society. The first concept addresses the issue of comparability of research results across studies. Subtle unmeasured differences may hamper the crude comparison of findings of different studies and call for sophisticated statistical methods. Once again, the book stresses that any choice of data analysis should be accompanied by a critical view and a deep understanding of the subject matter under investigation. Then, the last two chapters present a detailed account of strategies for the evaluation of the impact of a biostatistics research program in the society, which is the goal of modern scientific research.

Perugia, Italy
February 26, 2018

Elena Stanghellini

Preface

It almost never happened. My first introduction with Dr. Chiappelli resulted from a slight mishap in the administration's assignment of teaching assistants to lecturers. I received an email from him requesting that we meet before we began our working relationship for the upcoming semester. Being my first semester as a biostatistics teaching assistant, I was eager to showcase my passion for statistics within the health sciences. But before we even had the chance to meet, the error in our assignment had come to light—I received a follow-up email from him bidding me farewell, luck with whomever I would be assigned to, and left me with this: "Teaching is a very rewarding experience—hard work, but rewarding!"

This is but one of the many great pieces of advice that I would later receive from Dr. Chiappelli. By some miraculous change of events, I would not only remain assigned to his lecture that semester but also the semesters for the next year and half. During our first meeting, we quickly discovered our joint passion for biostatistics, research, and healthcare. As an ambitious premedical student, I was amazed by Dr. Chiappelli's pioneering work in convergent and translational science within modern healthcare. When I heard of his proliferative laboratory and research group, I knew that my once-in-a-lifetime opportunity was sitting across from me.

Fast-forward to the present—almost 3 years later—I have had the opportunity to publish in prestigious biomedical journals, be promoted to biostatistics lecturer, hold a senior laboratory position at UCLA, and now on track toward my dream—receiving a medical degree. None of this would have happened without the wisdom, guidance, and kindheartedness of Dr. Chiappelli. From biostatistics and research to the laboratory and medicine, the great deal of knowledge and experiences I have gained from him will certainly serve me in ways currently unfathomable.

And now our *first* book together! So, second to *The Omnipotent*, I thank you, Dr. Chiappelli, for this opportunity—I will cherish this book and the invaluable lessons you have taught me (and continue to teach me) for a lifetime. Your passion for teaching, advancing knowledge, and healthcare, has had a profound effect on me. I hope to one day pay forward your lessons to students of my own—and to think…it almost never happened.

I would like to thank our editor, Nicole Balenton, for her hard work and dedication to making this book perfect. Nicole is a brilliant young mind and former student who continues to surprise us with her excellence and ingenuity. I express my appreciation to my parents who have given me the utmost love throughout my entire life and my siblings, Arash and Angela, who I can always rely on for inspiration and wisdom. I thank Moses Farzan and Saman Simino for their continued support and friendship. Lastly, I extend my deepest gratitude and appreciation to Margaret Moore, Rekha Udaiyar, and the rest of the wonderful team at Springer for this opportunity and help throughout the process.

Haifa, Israel
January 2018

Allen M. Khakshooy

Acknowledgments

There is little that I can add to Dr. Khakshooy's excellent preface, except to thank him for the kind words, most of which I may not—in truth—deserve. This work is his primarily, and for me it has been a fulfilling delight to mentor and guide a junior colleague of as much value as Dr. Khakshooy's in his initial steps of publishing.

I join him in thanking Ms. Balenton, who will soon enter the nursing profession. Her indefatigable attention to detail and dedication to the research endeavors, and her superb and untiring help and assistance in the editorial process, has proffered incalculable value to our work.

I also join in thanking most warmly Ms. Margaret Moore, Editor, *Clinical Medicine*; Ms. Rekha Udaiyar, Project Coordinator; and their superb team at Springer for their guidance, encouragement, and patience in this endeavor.

I express my gratitude to the Division of Oral Biology and Medicine of the School of Dentistry at UCLA, where I have been given the opportunity to develop my work in this cutting-edge area of research and practice in healthcare, and to the Department of the Health Sciences at CSUN, where both Dr. Khakshooy and I have taught biostatistics for several years. I express my gratitude to the Fulbright Program, of which I am a proud alumnus, having been sent as a Fulbright Specialist to Brazil where I also taught biostatistics.

In closing, I dedicate this work, as all of my endeavors, to Olivia, who somehow always knows how to get the best out of me, to Fredi and Aymerica, without whom none of this would have been possible, and as in all, to only and most humbly serve and honor.

"… *la gloria di Colui che tutto move*
per l'universo penetra e risplende
in una parte più e meno altrove …"
(Dante Alighieri, 1265–1321; La Divina Commedia, Paradiso, I 1-3)

Los Angeles, CA, USA — Francesco Chiappelli

Contents

List of Videos

Part I

Fundamental Biostatistics for Translational Research

1 Introduction to Biostatistics

Contents

1.1 Core Concepts

Nicole Balenton

By the term "biostatistics," we mean the application of the field of probability and statistics to a wide range of topics that pertain to the biological sciences. We focus our discussion on the practical applications of fundamental biostatistics in the domain of healthcare, including experimental and clinical medicine, dentistry, and nursing.

As a branch of science, biostatistics encompasses the design of experiments, the monitoring of methodologies for appropriate sampling and accurate measurements, and the cogent analysis and inference of the findings obtained. These concerted activities are driven by the search of data-based answers to specific research questions. That is to say, biostatistics is the primary driver hypothesis-driven process by which research evidence is obtained, evaluated, and integrated in the growing knowledge-base of psychobiological processes in health and disease.

One strength of biostatistics lies in the unbiased nature of its inferences, which are based on the stringent laws of probabilities and bound by a rigid adherence to the requirements of randomness. Nonetheless, errors do occur in biostatistics, and the second area of strength of the field is its full awareness of these limitations. There are three types of errors possible in biostatistics: systematic errors, viz., mistake in planning and conducting the research protocol and in analyzing its outcomes; random errors, viz., mistakes that are consequential to situations and properties that occur randomly and are not under the control of the investigator (i.e., chance); and errors of judgment (i.e., fallacies), viz., making errors of interpretation rather than errors of facts.

This chapter discusses these fundamental concepts and introduces the timely and critical role of biostatistics in modern contemporary research in evidence-based, effectiveness-focused, and patient-centered healthcare. Emphasis is placed on the fact that there are, today, two principal approaches for looking at effectiveness: comparative effectiveness analysis, viz., comparing quantified measures of quality of life and related variables among several interventions, and comparative effectiveness research, viz., comparing

A. M. Khakshooy, F. Chiappelli, *Practical Biostatistics in Translational Healthcare*,
https://doi.org/10.1007/978-3-662-57437-9_1

several interventions in terms of relative differences in cost- and benefit effectiveness and in reduced risk. This chapter introduces these core concepts, which are explored in greater depth throughout this book.

1.2 The Scientific Method

Ask any lay person: "What is the scientific method?" and you will probably hear a response along the lines of "a method used in science." Be that as it may, it can be said with a degree of certainty that it is a method that almost every living human being has utilized—a bold statement, indeed, but the more we scrutinize its plausibility, the more we can consider its possibility.

A simple Internet search of "the scientific method" will produce millions upon millions of results that can provide anyone with an in-depth understanding of the scientific method. But that wouldn't be without hours of reading, critical analysis, surely a migraine or two, and of course the innate passion to learn. Now, what if there was a single word that could describe the scientific method as simply and accurately as possible—wouldn't that be of interest?

Why? This word, in and of itself, characterizes the curiosity, ingenuity, and advancement that is so particular to human nature. Asking questions like *Why? How? What?* and even *When?* and *Where?* are arguably the fundamental basis of this method we use in science. Granted, a small lie may have been told regarding just a *single* word. Rather, it can be said that there are a few single words that can just as simply and accurately achieve that, which is attempted to be imparted. So, let us refrain from claiming that there is a single word or many single words that can reflect what the scientific method is. Instead, it will be argued that it is the act of questioning, examination, or inquiry that lies at the heart of the scientific method.

Believe it or not, the scientific method was the scientific method before it was called the scientific method. As funny as that may sound, this meaning goes back to what was mentioned earlier in this chapter on the universal usage of the scientific method. Earlier, this may have not entirely resonated, but now with the "single word" description, that claim seems more conceivable. To stretch the argument further, a visit must be paid to the philosophers of old.

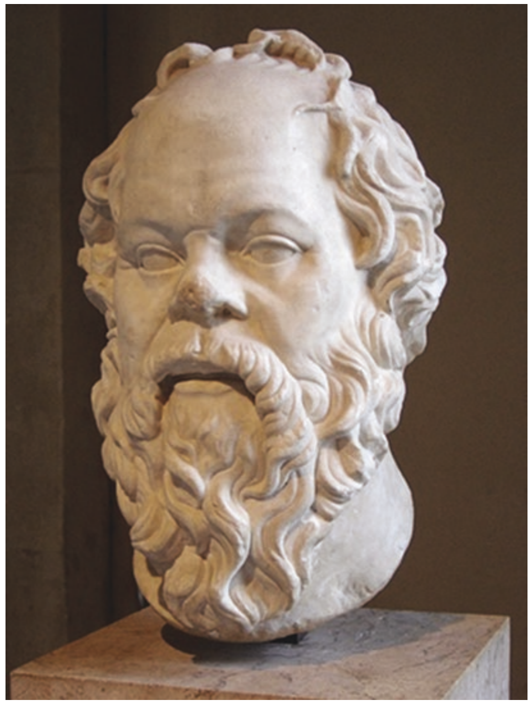

Fig. 1.1 A bust of Socrates in the Louvre (Gaba 2005)

Socrates (Fig. 1.1), regarded as the street philosopher of Athens, was infamous for soliciting the seemingly never-ending spiral of questions to those who passed by. Just as today, the people of ancient Greece considered it childish and aggravating when a man, uncanny to say the least, approached and perpetually probed them with odd and even embarrassing questions. Plato, his student, later dubbed this seemingly eerie behavior of Socrates as *elenchus*, today denoted as the Socratic method of inquiry. This method is one where, through a series of questions and investigations, one could attain the answers to a problem or, more philosophically, the truth of the matter.

Though the *truth* may seem out of scope or even unrelated to this subject matter, we shall see throughout this book that the understanding and

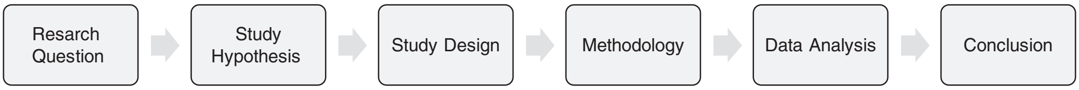

Fig. 1.2 Steps of the research process

attempt of attaining the truth may not seem so far-fetched after all.

Another large instrumental factor to our current scientific method was a grand-student of Socrates, Aristotle, a naturalist who explored the realms of logic, reasoning, and rationality that have largely influenced today's traditional Western thought. Of his many contributions, the techniques of inductive and deductive reasoning have played a large role in our scientific method today. We will return to this dichotomy of scientific reasoning later, but it must be noted that there currently exist many more influences to the evolution of our scientific method as we know it. On the same note, the scientific method still today is impartial to influences.

Finally, the scientific method is a method of investigating phenomena based on observations from the world around us, in which specific principles of reasoning are used in order to test hypotheses, create knowledge, and ultimately become one step closer in obtaining the truth. We must understand that there is no universal scientific method; rather there are fundamental concepts and principles that make this method of inquiry scientific. Moreover, the scientific method is ever-changing and ever-growing, such that the method itself is under its own scrutiny.

1.2.1 The Research Process

The research process can be argued to be the same as or a synonym for the scientific method. Though skeptics in this differentiation exist, for simplicity and practicality sake, we will distinguish the scientific method and the research process as the latter representing the actual application of the former.

The **research process** is a process that uses the scientific method to establish, confirm, and/or reaffirm certain pieces of knowledge supported by strong evidence or, as we may call it, proof. We use the research process to create theories, find solutions to problems, and even find problems to solutions we already have. In addition, the overarching goal of the research process is also an attempt to find some sort of truth. However, abstract this may seem, we can actualize its meaning by making the goal of the research process to be the culmination to an inference consensus or an ability to make a generalization of the whole based on its representative parts. Though the specific steps may differ based on their source, this book will take the steps of the research process as depicted in Fig. 1.2, along with a brief description of each provided in the following section.

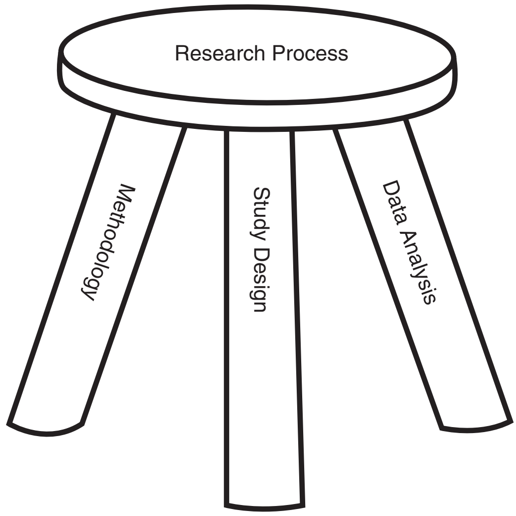

Fig. 1.3 Methodology, study design, and data analysis are the foundations of the research process

Lastly, the conceptualization of the research process as a whole can be interpreted to be a three-legged stool (Fig. 1.3) that sits on methodology, study design, and data analysis. This metaphoric description is crucial to the understanding of the research process such that each individual leg is equally valuable and important to the function of the stool. Just as the function of a stool is for one to sit, so too is the function of the research process: for one to gain otherwise unattainable knowledge. Hence, the integrity of the *stool* as a whole is placed in question should any single leg deteriorate.

1.2.1.1 Hypothesis-Driven Research

So, how does one begin the research process? The research process begins with nothing other than a question. The **research question**, simply put, is a question of interest to the investigator that serves as the driver of the entire process. The great value placed on this concept is an attempt to prove that the answer to this question is not only one that is interesting enough to warrant the need of a process but more importantly that the answer to it is both meaningful and useful. To take it many steps further, obtaining the answer to a research question could potentially prevent mass casualties in the world and help end world hunger.

Of course, this book is not a *how-to* manual on world peace. Rather, the premise that we are attempting to drive home is that not only can the successful answering of the research question be worthwhile but that we may very well not always be successful in obtaining an answer. Thus, research questions are chosen based on a certain criterion easily remembered as the acronym FINER. We say that a research question must be: feasible, interesting, novel, ethical, and relevant. Though there can be a never-ending list of categories of research questions (Table 1.1), below we provide a few types of research questions that are relevant to our specific interest in this book[1].

Table 1.1 Types of research questions

Types of research questions
Descriptive—attempts to simply describe that which is occurring or that which exists
Relational—seeks to establish, or to test the establishment of, a specific relationship or association among variables within groups
Causal—developed to establish a direct cause-and-effect relationship either by means of a comparison or by means of a prediction
PICO(TS)—describes specific criteria of research as they refer to the patient(s), the interventions, and its comparators that are under consideration for a given sought outcome, under a specified timeline and in the context of a specific clinical setting

A hypothesis, commonly referred to as an educated guess, is seen as both a starting point and guiding tool of the research process. But, was it not mentioned earlier that it is the research question that is the starting point? Indeed! Here is where the intimate relationship between the research question and the study hypothesis is made clear. The **study hypothesis** is nothing more than the research question stated positively (i.e., research question is changed from question format to statement format.) The disparate forms of hypotheses are further discussed in *Hypothesis Testing* in Chap. 5.

The **study design** serves as the infrastructure or the system we create that aids in answering the research question. The design of any research process is, obviously, dependent on both the peripheral and inherent details of the research question like the specific population, disease, and therapy that is being studied.

The **methodology** of the research process is concerned with the process of measuring and collecting the necessary information (which we call data, discussed further in Chap. 3) regarding the specific population of interest depicted in the research question. As further elaborated in Chap. 3, because it is seemingly impossible to comprehensively study an entire population, we obtain data from a sample that is a representative of the entire population that can survive this comparison.

Data analysis is the statistical techniques and reasoning tools utilized in the examination of the collected information, i.e., data. Some have regarded this section as the results of the study, in which the evidence obtained is used in hopes of proving or disproving the conjectured hypotheses.

Lastly, the **conclusion** is the researcher's attempt to answer the research question relative to the results that were obtained. It is at this point that our initial concepts of inference consensus and truth determination converge. Though the

[1] Note the acronym stands originally for population, intervention, comparator, outcome, timeline, and setting; the latter two are parenthetic such that they are not always used or available to use; in any case they can be described as PICO, PICOT, or PICOS research questions.

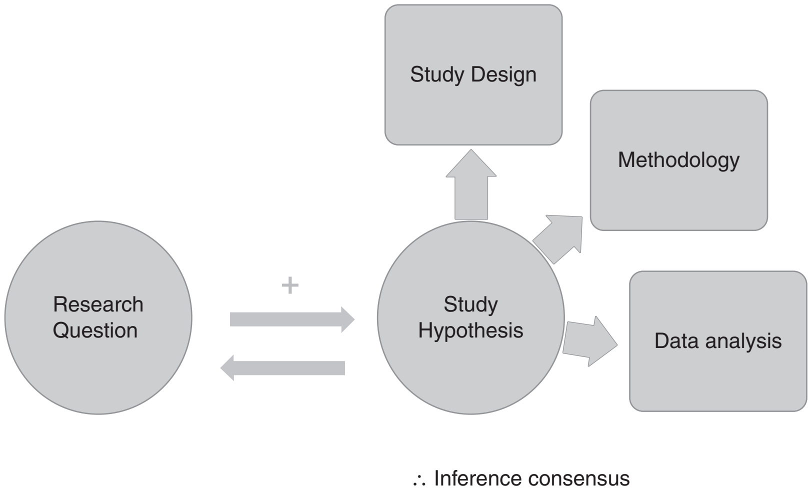

Fig. 1.4 The study hypothesis is the driving force of the research process, hence hypothesis-driven research

analysis of the data is meant to provide some sort of concrete evidence to influence the decision-making process on behalf of the postulates, it is unfortunately not that forthright. Statistical analysis allows us to put limits on our uncertainty regarding the issue at hand, but what it does not clearly allow is the absolute proof of anything. Thus, when arriving at the conclusion of a study, the results are unable to provide an absolute truth statement when all is considered. Rather, its application is more practical in disqualifying substantiated claims or premises.

Similar to the fundamental principle in the US Justice System of "innocent until proven guilty," so too exists a principle that is central to the scientific method and the research process in regard to the treatment of hypotheses within a research study. We commonly retain a basic hypothesis of the research (namely, the null hypothesis discussed in Chap. 5) such that we cannot adequately prove its absolute truth for obvious reasons. Instead, what we *are* capable of is proving its absolute falsehood. Subsequently, the pragmatism that is intrinsic to our conclusion is the ability to make an inference. Upon evaluation of the results, an inference is made onto the population based on the information gleaned from the sample.

A quick glance at the crude descriptions of each step of the research process shows the impact of the research question along the way. Then, after equating the research question with the study hypothesis, it can now be understood why the research process is referred to as *hypothesis*-driven research (Fig. 1.4). It is the study hypothesis that is the driver of all three legs of the stool (methodology, study design, and data analysis), which culminate into the making of a potential inference.

1.2.1.2 Errors in Research

Statistics in translational healthcare pervades the scientific literature: its aim to improve the reliability and validity of the findings from translational research. As we progress toward a more technologically advanced future with greater accessibility, it seems as though we are constantly bombarded with so-called *proven* research findings, medical breakthroughs, and secretive therapies on a daily basis. It even goes as far as having distinct research findings that directly contradict one another! Recently, we have witnessed a prevalence in the retraction of research papers that, just a few years earlier, were highly regarded as pivotal to modern-day science. Though the majority of retracted papers stem from ethical concerns, there are papers that have so-called "fudged" the numbers or simply have improperly handled the statistics. These mishandlings also stretch beyond just the statistics, which we categorize as errors.

Naturally, the susceptibility of the research (and the researcher) to error is inevitable. The effect of errors is most felt during the determination of results, or more specifically when establishing statistical significance. Discussed in more depth in

Chap. 5, the establishment of statistical significance (or lack thereof) is an imperative and necessary step in the substantiation of our results (i.e., when moving from data analysis to conclusion). This lends a hand to the importance placed on inherent errors and introduced biases that are, unfortunately, contained in many published research today.

Just as the research process is a three-legged stool, so too is the establishment of statistical significance (Fig. 1.5). The process of obtaining statistical significance sits on three forms of error: systematic errors, errors of judgment (i.e. fallacies), and random errors. We do not have the full capability of understanding the intricacies of each error just yet, but for the moment, it is worth attempting to briefly describe each one.

Systematic errors are just as they sound—errors in the *system* we have chosen to use in our research. What systems are we referring to? That would be the study design. Erroneously choosing one design over another can lead to the collapse of our ultimate goal of attaining statistical significance. Luckily, systematic errors are one of the few errors that have the ability of being avoided. We can avoid systematic errors by simply selecting the best possible study design. There are many factors that lead to the appropriate selection of a study design like the type of research question, the nature of the data we are working with, and the goal of our study to list a few. But more importantly, the risk of running a systematic error (choosing a poor study design) is that it will always produce wrong results of the study.

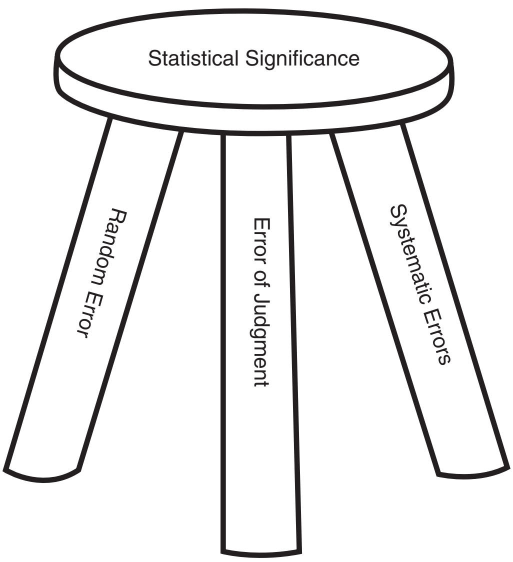

Fig. 1.5 The three basic types of error that mediate statistical significance

The second type of error are **errors of judgment** or **fallacies**. To elaborate, these are errors that are grounded in biases and/or false reasoning (i.e., a fallacy), in which the improper use of logic or rational leads to errors in scientific reasoning. It can be argued that these are the most dangerous errors in research as they are subjective to the researcher(s). In Table 1.2, we provide a list of the various types of fallacies.

Table 1.2 A description of several common types of fallacies or biases that may occur in scientific reasoning related to research

Errors of judgment/fallacies
Hindsight bias (“knew-it-all-along” effect): The foretelling of results on the basis of the previously known outcomes and observations; subsequently testing a hypothesis to confirm the prediction to be correct. For example, taking it for granted that the Sun will not rise tomorrow
Recomposing-the-whole bias (fallacy of composition): The bias of inferring a certain truth about the whole, simply because it is true of its parts. For example, since atoms are not alive, then nothing made up of atoms is alive
Ecological inference bias (ecological fallacy): The act of interpreting statistical data (i.e., making statistical inferences) where deductions about the nature of individuals are made based on the groups to which they belong. For example, America is regarded as the most obese country in the world today; therefore my American cousin who I’ve never met must be obese!
Fallacia ad hominem (**shooting the messenger**): The fallacy of blaming a poor outcome on the fault of others. For example, “It’s not my fault the results were bad, it’s the fault of the engineer of the statistical software!”
Fallacia ad populum et ad verecundiam (**“everybody does it!”**): The fallacy of common practice or of authoritative knowledge. For example, “I just did it the way everybody else does it!” or “This is how my Principal Investigator does it!”
***Fallacia ad ignorantiam et non sequitur* (“*Just because!*”):** The fallacy of common practice without any certain proof that what is done is appropriate. For example, **“**I did it this way because I don’t know of a better way, that’s just how I learned to do it!”

The third type of errors in research are **random errors** which can arguably be the most common of the bunch. These are errors that are essentially beyond control—meaning that no matter what, this type of error cannot be avoided or prevented in entirety. Better yet, we can be certain of its occurrence simply because we (the researcher, study subjects, etc.) are all human and error is imbedded in our nature.

Although this should not be as alarming as its doomsday, description makes it to be. Why? Because statistics are here to save the day! One of the primary functions of the statistical tools and techniques later described in this book is to decrease or *fractionate* random error, thereby minimizing its potentially detrimental effects on our results. On the other hand, the presence of error in our study can also serve a positive purpose in so far as it takes into consideration the individual differences of the study subjects. Truthfully, there can be an entire field within statistics dedicated to the process of and value behind the minimization of error. For now, we can assure that its presence will be felt in the majority of the sections that follow in this book.

1.2.2 Biostatistics Today

Biostatistics—the word itself may seem intimidating at first. Should one want to impress their friends and family, mentioning you are studying biostatistics is an easy way to accomplish that. But however intimidating the word may seem, the actual study of biostatistics should not be feared. Moreover, the roots of the word may hint at its actual concept and study: *bio* and *statistics*. Hence, a layperson may perceive the study of biostatistics to mean the statistics in biology or life statistics, a weak interpretation of the word.

Although we may use biostatistics in the field of biology, the more representative meaning that we will side with is the latter—namely, the statistics behind human life. Further, biostatistics is a specific branch of statistics that utilizes information that is particular to living organisms. But it must be made clear that the fundamental tools and concepts of biostatistics are no different than those of statistics itself. Moreover, it is the overarching theme and ultimate purpose behind the utilization of these techniques that make it specific to *bio*statistics.

The study of biostatistics is not limited to any one field, like biology. One of the great virtues of this study is that it involves a multidisciplinary collaboration between the wealth of today's studies that have to do with human life. To name just a few, these disciplines range from psychology and sociology to public health and epidemiology and even to medicine and dentistry.

Thus, the utilization of biostatistics today is the application and development of statistical theory to real-world issues particular to life as we know it. Additionally, the aim we are working toward is solving some of these problems, in hopes of improving life as a whole. So, we can see how it would not be uncommon to hear the biomedical sciences as being the broad discipline subjective to biostatistics. But since the nature of this book is pragmatism, we will simplify its comprehensive discipline from the biomedical sciences to the health sciences. Hence, taken together, biostatistics lies at the heart of the research process in the health sciences.

1.2.2.1 Relevance for the Health Sciences

The health sciences are composed of a large variety of applied scientific fields that pertain to the usage of science, mathematics, technology, and engineering in the delivery of healthcare and its constituents. The health sciences cover a wide variety of disciplines that are not solely limited to conventional Western medicine; rather they stretch to both traditional and alternative medical modalities. That being said, it is not so much the actual practices of these medical modalities that are of concern in this book; rather it is the methods of utilization of the collected information from these practices that is of chief concern.

When we bring biostatistics into the conversation, we see that its introduction to the health sciences serves the purpose of our research. Just as we spoke of the importance of the research question, it is the health science-based research question that requires biostatistical theory to be

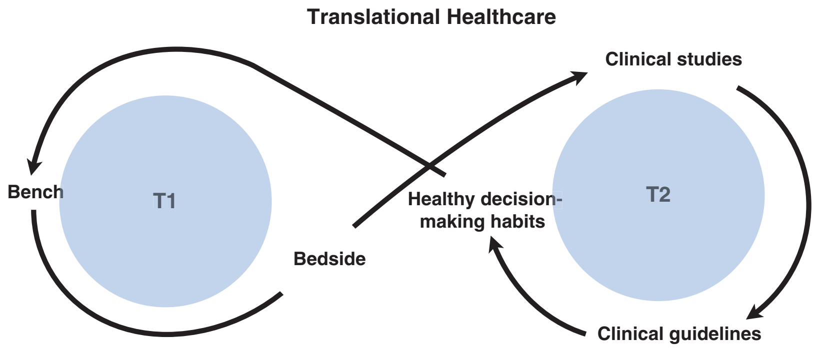

Fig. 1.6 Translational healthcare model

answered. Moreover, we can now perceive the value of the hopeful answer that we obtain from the health science-based research question. The significance of this answer being that it is the best possible answer to a problem that seeks the betterment of both the healthcare field and its constituents.[2]

Conclusively, a primary aim of this book is to provide an understanding of the basic principles that underlie research design, data analysis, and the interpretation of results in order to enable the reader to carry out a wide range of statistical analyses. The emphasis is firmly on the practical aspects and applications of the methodology, design, and analysis of research in the science behind translational healthcare.

1.2.2.2 Research in Translational Healthcare

A biostatistics course is essential, if not mandatory, to a student in the health sciences. This is mainly for the acquisition of basic and scientific statistical knowledge that pertains to the specific area that is being studied within the health sciences. But as we progress from today's students to tomorrow's professionals, the great value of biostatistics arises within the field of translational healthcare.

As this is being written, the fate of US healthcare, for better or worse, is uncertain, but what is certain is the direction that the field is moving toward as a whole: placing focus on the individual patient. The field of translational healthcare is one which takes a patient-centered approach that translates health information gained from a research setting to the individual patient and, if effective, translated to benefit all patients. Furthermore, this is the crude conceptualization of the two primary constructs of the science of translation (or translational medicine as it was first introduced)—namely, translational research and translational effectiveness.

In theory, **translational research** refers to the first construct (T1) and **translational effectiveness** as the second construct (T2), in which this book has been divided as such accordingly (Fig. 1.6). The first half of this book is responsible for expounding on the fundamentals of translational research and its practical application in healthcare, such that the methods to be discussed aid in the translation of information from "bench to bedside." This is essentially the process of going from the patient to the laboratory bench and back to the patient. Namely, new knowledge of disease mechanisms gained from the laboratory bench is transferred to the development of new methods for diagnosis, therapy, and prevention that directly benefit the patient.

[2] For example, just a few years ago citizens of the United States questioned the lack of universal healthcare in their country. This was deemed as a problem to the overall well-being of the United States and its constituents, which was supported by epidemiological evidence among others (i.e., mortality rates, prevalence of preventable diseases, etc.). Moreover, the evidence proved that there was much need for an affordable and accessible healthcare plan that would solve the problems that resulted from a lack of universal healthcare in the United States. Hence, in 2008, the US Congress passed the Affordable Care Act which was aimed at settling this real-world problem for the overall well-being of the healthcare field (i.e., legislative policy) and its constituents (i.e., US citizens).

On the other hand, the succeeding half is responsible for the introduction of the second construct of translational science, namely, translational effectiveness. This is referred to as "result translation," in which the results that are gathered from clinical studies are translated or transferred to everyday clinical practices and healthy decision-making habits. Although we have bisected the two based on their distinct purposes, methods, and results, both enterprises coalesce to the ultimate goal of new and improved means of individualized patient-centered care.

In brief, the most timely and critical role of biostatistics in modern contemporary research in healthcare today appears to be in the context of:

(a) Establishing and evaluating the best available evidence, in order to ensure evidence-based interventions
(b) Distinguishing between comparative effectiveness analysis, which is designed to compare quantified measures of quality of life and related variables among several interventions, and comparative effectiveness research, which aims at comparing several interventions in terms of relative differences in cost- and benefit effectiveness and in reduced risk, in order to ensure effectiveness-focused interventions
(c) Characterizing novel biostatistical toolkits that permit the assessment, analysis, and inferences on individual, rather than group, data to ensure the optimization of patient-centered interventions

1.2.3 Self-Study: Practice Problems

1. How does the process of using the scientific method begin?
2. List and provide a brief description of the steps of the research process?
3. What are the legs that represent the stool that is the research process? What happens if one of the legs is compromised?
4. What is the difference between the research question and the study hypothesis?
5. True or False: The best type of research study is one that can conclude the absolute truth.
6. What are the legs that represent the stool that is statistical significance? What happens if one of the legs is compromised?
7. Which of the three most common types of errors are avoidable? Which are unavoidable?
8. You have just finished taking your first biostatistics exam and are unsure how well you performed. Later that week, you receive your results and see that you received an A—and exclaim: "I knew I was going to ace that!"—Which of the biases was taken advantage of in this scenario?
9. True or False: All forms of error introduced during the research process negatively impact the study as a whole.
10. Translational healthcare is comprised of two enterprises. What are these two enterprises and what does each represent?

(*See back of book for answers to Chapter Practice Problems*)

2 Study Design

Contents

2.1 Core Concepts

Nicole Balenton

The composition of the basic principles that act as the foundation of the research process is conceptualized as a three-legged stool. This chapter highlights the first of the three legs of the stool—namely, study design—that acts as the blueprint for researchers to collect, measure, and analyze the data of their health topic of interest. The study design hinges on the research topic of choice.

As the backbone of any successful scientific research, the study design is the researcher's strategy in choosing various components of a study deemed necessary to integrate in a coherent manner in order to answer the research question. The design chosen affects both the results and the manner in which one analyzes the findings. By obtaining valid and reliable results, this ensures that the researchers are able to effectively address the health research problem and apply the findings to those most in need.

The success of any scientific research endeavor is established by the structure of the study design, offering direction and systematization to the research that assists in ultimately understanding the health phenomenon. There are a variety of study design classifications; this chapter primarily focuses on the two main types: diagnostic studies and prognostic studies. We further explore their respective subcategories and their relation to scientific research in translational healthcare.

2.2 Conceptual Introduction

As one of the three fundamental pillars of the research process, the design of a research study is essentially the plan that is used and the system

A. M. Khakshooy, F. Chiappelli, *Practical Biostatistics in Translational Healthcare*,
https://doi.org/10.1007/978-3-662-57437-9_2

employed by the researcher. The specific organization is subjective to the particulars of the object of the study. The "particulars" we mention are pieces of information that refer to things (i.e., variables, discussed further in Chap. 3) such as the population and the outcome(s) of interest that are being studied. We can conceptualize the study design to be the infrastructure or the organization of the study that serves the ultimate goal of the research that is being done.

Let's say, for example, that you and your significant other decide to build your dream home together and we will largely assume that you have both also agreed on the final plan of the whole house, i.e., where each room will be and their individual uses. But as any contractor will tell you, the foundation of the house is of utmost importance because it sets the precedent for the building as a whole. Sure, it is pertinent to have a plan of what each room should be, but through the creation of the foundation is where the proper piping, plumbing, and electrical groundwork are set for the ultimate *design* of each room. This is exactly the purpose and relationship between a study and its design.

As researchers, we must be fully cognizant of the intricate details of our study in order to create a study design that can facilitate the goal of our research. Luckily, there is no creation necessary on our part as there are a multitude of study designs we can select for the specific components of our study. Moreover, correctly selecting the infrastructure at the outset is an early step in preventing faulty outcomes during the research process (i.e., preventing systematic errors).

Now, if your dream home consists of a two-story building, would you build a foundation, create a blueprint, and buy the accessories necessary to build a one-story house? Or if the zoning regulations of the property prohibited the building of a multistory house, could you still move forward with the successful materialization of the dream? Of course not, these would be disastrous! Similarly, we would not dare to, on purpose, select a study design that is specific to one method of research when our own research is concerned with another.

This chapter discusses distinct and comprehensive study designs relative to scientific research. More specifically, the designs and their subcategories are tailored toward the infrastructure that is necessary for research in translational healthcare. The summarizing schematic of the disparate study designs is shown in Fig. 2.1.

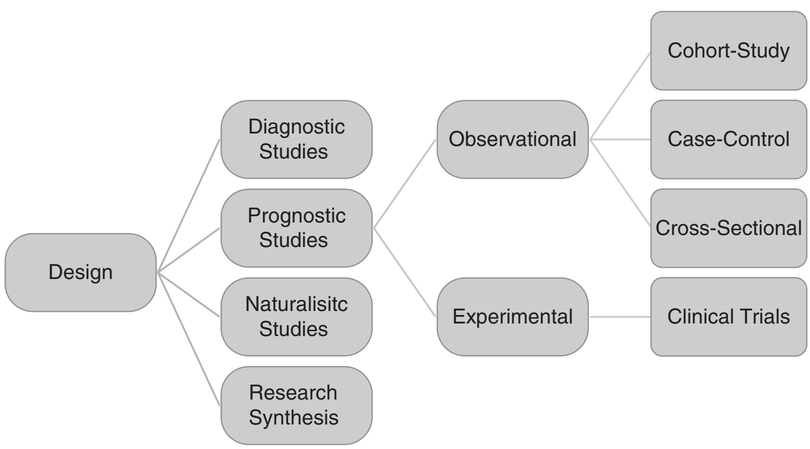

Fig. 2.1 The various study types including observational studies, experimental studies, naturalistic studies (Naturalistic study, often referred to as qualitative, participant observation, or field research design, is a type of study design that seeks to investigate personal experiences within the context of social environments and phenomena. Here, the researcher observes and records some behavior or phenomenon (usually longitudinally) in a natural setting (see Chiappelli 2014), and research syntheses (Research synthesis—a type of study design that utilizes a PICOTS research question (see Chap. 1, Sect. 1.2.1.1), in which relevant research literature (the sample) is gathered, analyzed, and synthesized into a generalization regarding the evidence—this is the fundamental design behind evidence-based healthcare (see Chiappelli 2014)