Pedro J. Gutiérrez Diez · Irma H. Russo
Jose Russo

The Evolution of the Use of Mathematics in Cancer Research

Springer

The Evolution of the Use of Mathematics in Cancer Research

Pedro J. Gutiérrez Diez • Irma H. Russo • Jose Russo

The Evolution of the Use of Mathematics in Cancer Research

Springer

Pedro J. Gutiérrez Diez
University of Valladolid
Valladolid, Spain
pedrojos@fae.uva.es

Irma H. Russo
Fox Chase Cancer Center
Philadelphia, PA 19111, USA
irma.russo@fccc.edu

Jose Russo
Fox Chase Cancer Center
Philadelphia, PA 19111, USA
jose.russo@fccc.edu

ISBN 978-1-4614-2396-6 e-ISBN 978-1-4614-2397-3
DOI 10.1007/978-1-4614-2397-3
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2012930389

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To our parents, who through their efforts, lessons and understanding forged in us a sense of accomplishment and greatness.

PJGD, IHR, JR.

To my wife Araceli, my support, so wonderfully rational, so wonderfully emotive.

PJGD

To all our trainees and students, from whom we have received more than they ever expected to give us.

PJGD, IHR, JR.

Preface

This book provides an exhaustive and clear explanation of how statistics and mathematics have been used in cancer research, and seeks to help advanced students of biostatistics and biomathematics as well as cancer researchers to achieve their objectives. To do so, state-of-the-art biostatistics and biomathematics are described and discussed in detail through illustrative and capital examples taken from cancer research work already published. The crossed examination of the statistical, mathematical and computational issues arising from the selected examples redounds to a didactic, homogeneous and unified vision of the application of statistics and mathematics in biomedicine, especially to the study of cancer, and illustrates the capability of these logical sciences in biomedical research. As a result, the book provides a guide for cancer researchers in using statistics and mathematics, clarifying the contribution of these logical sciences to the study of cancer, thoroughly explaining their procedures and methods, and providing criteria to their appropriate use.

Indeed, this book is designed for advanced students and researchers pursuing the use of biostatistics and biomathematics in their investigations and research in biology and medicine in general, and in cancer in particular. The main virtue of the book is the follow-through that is available by reading the different examples, in a relevant and timely reading that facilitates the understanding of the key aspects underlying the applications of statistics and mathematics in biomedicine, and that provides complete coverage of the most relevant issues in biostatistics and biomathematics. Each chapter has been conceived as a part in the whole in such a way that information flows easily, on the one hand explaining in a concise and clear way a particular subject, and on the other connecting its results with those in the previous and following chapters. Thanks to the use of selected and relevant examples taken from the scientific literature on cancer research, the result is a self-contained book on medicine, statistics and mathematics, which illustrates the potential of biostatistics and biomedicine in biomedical research. Focusing on the achievements that biostatistics and biomathematics have already obtained, researchers can perceive the high returns that the use of statistics and mathematics yield in cancer research, and thanks to the detailed discussion of the applied statistical and mathematical techniques, they can deduce the criteria and motif for finding the appropriate use of these formal disciplines.

The primary audience of the book is advanced undergraduate students and graduate students in medicine and biology, and cancer researchers who seek to learn how statistics and mathematics can help in their future research. We assume no advanced knowledge of statistics and mathematics beyond the undergraduate level. However,

the reader should have a minimum formation in these disciplines and be familiar with the contents of undergraduate textbooks on mathematical analysis and biostatistics.

The use of statistics and mathematics in biology and medicine is today increasing, and already forms part of the core of both theoretical and empirical biomedical research. We hope with this book to contribute to a better comprehension of the procedures, methods, criteria and applications of biostatistics and biomathematics in medicine, especially in cancer research.

The authors

Acknowledgements

Our special acknowledgement and thanks to Mr. Alan Hynds, B.A., and Ms. Patricia A. Russo, M.F.A., for their insightful style suggestions and editing assistance, and to *Pathology Consultation Services* from Rydal, PA, that has financed the writing and editing of this book.

Pedro J. Gutiérrez Diez is also grateful to the helpful suggestions and comments of Dr. Luis Borge and to the financial support from Education and Science Department, Spanish Government, research project ECO2009-10231, and from Education Department, Castilla and León Autonomous Government, research project VA016A10-1.

Contents

Chapter 1
Historical Introduction

Abstract This chapter summarizes the different ways in which statistics and mathematics have been applied to biosciences. Following chronological and historical criteria, the different stages describing the evolution of biostatistics and biomathematics are succinctly explained, quoting the main contributors and their work, and specifying the particularities of each of these scientific disciplines and the links between the two.

1.1 Biomedical Sciences and Logical Sciences

To provide a definition of Science including all its relevant attributes is almost impossible. However, to our purposes, Science can be defined as a method to obtain knowledge. In fact, the main characteristic of Science is not the obtained knowledge, but the particular manner in which it is obtained. When a palmist "reads" your palm and affirms that you are ill and a physician explores you and arrives at the same conclusion, if indeed you are ill, the two prognoses are not the same or equivalent despite the fact that you are ill. Unlike the palmist's prediction, the conclusion of the physician is scientific, and this scientific characteristic is due to the particular method followed by the medic, different from that applied by the palmist. As a matter of fact, knowledge is described as scientific if it has been obtained applying a particular method, the scientific method, which constitutes the core of Science. This specific procedure to get knowledge is based on the elaboration and contrast of theories, the basic units of scientific knowledge. In Science, manner and content are deeply and intrinsically related, since the elaboration of theories is not only the method to obtain knowledge but also constitutes the materialization of such knowledge. This is why, at the beginning of this paragraph, Science was defined as a particular method to obtain knowledge.

In essence, a theory is a set of hypotheses from which, following logic and formal reasonings, some conclusions are derived. These conclusions, called "theoretical implications" for obvious reasons, are contrasted with reality, and this contrast determines the acceptance or rejection of the theory. If the confrontation of the theoretical conclusions versus the reality is good enough, the theory is accepted; if, on the contrary, the reality is not explained by the theoretical implications, the theory is rejected. Since all the implications directly derive from the hypotheses, if a theory

P. J. Gutiérrez Diez et al., *The Evolution of the Use of Mathematics in Cancer Research,*
DOI 10.1007/978-1-4614-2397-3_1,

does not work, it is only necessary to modify the set of hypotheses, by adding, removing or changing some of them. Then, as explained before, the theory is not only the method to obtain knowledge but also constitutes the unit of such knowledge. By its own nature, the theory is in essence an open corpus of knowledge. A theory is open because it must be contrastable with reality, that is, it must be susceptible to rejection and improvement as well as acceptance. As a consequence, Science advances thanks not only to the elaboration of new theories but also to the rejection of the old ones.

This scientific method is common to all sciences, something that makes the transmission and sharing of knowledge between disciplines possible. These flows of results across scientific fields have always proven to be very fruitful. Indeed, today, multidisciplinary and transdisciplinary analyses are crucial for scientific progress, increase over time, and originate new scientific branches born from two or more previously existing scientific fields. This is the case of biomathematics and biostatistics, the disciplines being studied in this book. They are a consequence of the application of mathematics and statistics to the analysis of biological and medical phenomena, or, alternatively, of the necessity of medicine and biology to count on mathematical and statistical results, analyses and techniques.

How did logical and bio-medical sciences join? In principle, biosciences and logical sciences are not very close scientific fields. Biosciences apply the scientific method to study biological and medical questions. With the goal of satisfactorily explaining a biomedical phenomenon, a biological or medical theory starts from a set of hypotheses that describe the analyzed reality from a new or different perspective. From these hypotheses, carrying out and applying procedures, experiments and reasonings involving well established and accepted biological, medical, physical and chemical theories and results, the proposed theory reaches some theoretical conclusions. These implications, directly originated in the hypotheses, are contrasted with the reality. As explained above, if the confrontation is good enough, the new explanation of the biomedical phenomenon is accepted; if not, the theory is rejected and its hypotheses revised. It is worth noting that when designing experiments and extracting conclusions from the hypotheses, and when putting these theoretical conclusions to the test, accepted and well established results and theories coming not only from biology and medicine but also from physics and chemistry are used and applied. As stated before, the flow of results across scientific fields has always been very fruitful, and, indeed, are indispensable for the advance of Science.

Logical sciences, on the other hand, deal with logical and formal questions. The analyzed reality is not of a biological, medical, physical or chemical nature, but logical and formal. Then, unlike natural and biological science theories, which must originate implications in agreement with an external reality, logical science theories must be consistent and coherent with the previously accepted logical science theories: the contrast and confrontation of the theoretical implications of a logical science theory is not of an external but an internal nature. However, leaving aside this fact,

the procedure to obtain knowledge is the same for logical sciences as for natural and biological sciences: to elaborate and to test theories[1].

As commented above, transmission and sharing of knowledge between scientific disciplines has always proven to be crucial for scientific advance. These flows of results appear between natural and biological sciences—for instance, medicine usually appeals to biology, chemistry and physics—, between logical sciences—statistics makes use of mathematical results—, and also across all these disciplines—logical sciences have proven to be very useful in nature sciences, and, indeed, physics or chemistry is unimaginable without applying mathematics or statistics. And what about the application of logical sciences to bio-medical sciences? Without any doubt, it can be asserted that, today, advances in biology and medicine require the use of mathematics and statistics.

Indeed, after being progressively applied to physics (15th century), chemistry (17th century), engineering (18th century) and economics (19th century), logical sciences (firstly statistics and afterward mathematics) have begun to be used for the analysis of biological questions. Indeed, as a result of the ability of modern logical sciences to describe complex and interrelated behaviors, the application of mathematical and statistical models to study biological and medical phenomena increases over time.

Leaving aside purely exogenous factors, the main characteristic of biomedical phenomena is the existence of numerous and complex relationships between entities (cells, bacteria, human organs, genes, living beings, species, ...). To analyze biomedical questions, logical sciences were firstly applied to quantify these relationships and to obtain the correlation between observed biological behaviors. The development of statistics in the 19th century made the description—but not the explanation—of the relationships and correlations between behaviors of biological entities possible. The statistical analysis of the biological behaviors soon proved its ability to specify the relationships between behaviors, the cause-effect directions, the existence of clusters, etc., and as a result, biostatistics is today a basic tool in medicine and biology.

Once the interrelationships were statistically described, the next step was to explain their origin. Hand in hand with medical and biological experimentation, the use of systems of equations in the 20th century helped to elucidate why the particular interrelationships between bio-entities appeared. The main virtue of the systems of equations is their capability to explicitly state such interrelationships, therefore providing a first explanation of the interrelated behaviors. Indeed, most modern biomathematics relies on this kind of mathematical analysis, and, in particular, the use of systems of difference or differential equations is today the more frequent mathematical technique to explain interrelated biological behaviors.

Being able to count on a system of equations describing, detailing and elucidating the biological phenomena and the interrelationships between the involved

[1] Actually, all sciences must be internally and externally coherent and consistent. In natural and biological sciences the bias is towards external consistency, while in logical sciences it is to internal consistency.

bio-entities soon opened up the possibility of controlling these behaviors, something particularly interesting when designing therapies. Born in the 1950s and originally designed to solve economic and engineering problems, optimal control theory provided the mathematical tools to tackle this question. The design of optimal therapies through the control of the interrelationships between bio-entities that appear in a particular therapy, described by the system of equations, is today an expanding area of biomathematics, but is not the only application of optimal control theory. Indeed, this mathematical technique can also be applied to obtain a more complete explanation of the biological and medical phenomena than those provided by the use of systems of equations[2].

Although this book will not analyze bioinformatics or computational biology questions, it is worth noting that the development of biostatistics and biomathematics would have been impossible without the help of informatics. To carry out the sophisticated statistical analyses that characterizes today's biostatistics, and to solve the systems of equations and the optimal control problems used by biomathematics, require the implementation of complex computational procedures. Indeed, at the present time, there are a great number of algorithms, programs and packages specifically designed for biostatisticians and biomathematicians. However, bioinformatics is much more than this, since it also deals with the elaboration of computational procedures and computational hardware for simulating and replicating biological behaviors. This is the case of such fields as artificial intelligence, artificial neural networks and evolutive robotics, with increasing importance in bioinformatics.

All these questions will be developed and explained in detail in the following chapters of this book. After this introductory chapter, devoted to briefly explaining the nature and history of biostatistics and biomathematics, the applications, techniques and current state-of-the-art of these scientific disciplines will be described. Chapters 2–4 will focus on biostatistics, respectively on descriptive biostatistics—Chap. 2—and inferential biostatistics—Chaps. 3 and 4. Chapter 5 will be devoted to equations, and Chaps. 6 and 7 will analyze the use of systems of equations. Finally, optimal control will be analyzed in Chaps. 8 and 9, being game theory studied and commented on in Chap. 10. Whenever possible, our discussion will be related to the study of cancer.

Before commencing our analysis and examination of biostatistics and biomathematics, let us define and contour these terms. In this respect, although statistics is a branch of mathematics, we will differentiate between these two terms, as many authors do. In particular and for our purposes, we will consider statistics as the numerical and logical analysis of non totally predictable phenomena, and we will define mathematics as the analytical and logical analysis of predictable phenomena.

---

[2] See Rocklin and Oster (1976), Perelson et al. (1976, 1978), Perelson et al. (1980), and Gutiérrez et al. (2009).

1.2 Biostatistics: Historical Notes

As a science, statistics can be defined as the scientific study of the numerical data that emerge from non totally predictable phenomena. All the parts in this definition are important. Firstly, statistics is a science, and therefore it applies the scientific method described in Sect. 1. Secondly, statistics only considers numerical data, ignoring either the nature or the intrinsic logic underlying the analyzed phenomena. And thirdly, the numerical data must not be totally predictable, that is, the analyzed phenomena must involve some degree of uncertainty about their results.

It is obvious that, for several reasons, numerical data arising from biomedical phenomena are not totally predictable. In biology and medicine, most phenomena are affected by a great number of causal factors, and some of them are uncontrollable or simply unidentifiable. In addition, even assuming that the causal factors underlying such phenomena were totally identifiable and controllable, biomedical magnitudes and variables are not always exactly measurable, and the mere existence of measurement errors imply uncertainty about the arising numerical data. Therefore, it is not odd that statistics were soon applied to study problems of biology and medicine[3]. Indeed, although in its origin during the second half of the 17th century statistics was applied to analyze demographic and economic questions and to study chance games, the use of statistics in problems of biology and medicine developed relatively early at the beginning of the 19th century, only preceded by its application to astronomy during the second half of the 18th century. As a fact, the first biostatistics analyses were done by Adolphe Quételet in 1832 after meeting his contemporaneous mathematicians and astronomers Joseph Fourier, Siméon Poisson and Pierre Laplace. These extraordinary scientists had previously applied statistics to astronomy, and transmitted to Quételet the interest in statistics. The result was the application by Quételet of the theory and practical methods of statistics to analyze the physical characteristics of man, not only by considering ratios as an important statistical tool—the body mass index was created by Quételet—but also by doing cross-section analyses and statistical characterizations for the data distributions of physical attributes of humans.

Quételet's (1832, 1835, 1842) pioneering work on biostatistics was continued by Francis Galton [1822–1922], who applied the data distribution analysis initiated by Quételet to the study of heredity. Despite the fact that Quételet's studies constitute the first biostatistic analyses, it is Galton who is considered the father of biostatistics for two main reasons. First, his methodology became the basis and foundation for the application of statistics to medicine and biology, and second, Galton's work opened up the research avenue followed by biostatistics until the first half of the 20th century, centered on the reconciliation between the evolutive mechanisms of natural selection and mendelian genetics.

Regarding the methodological contributions of Galton, he was the inventor of paramount statistical techniques and concepts such as standard deviation, correlation

[3] The use of mathematics other than statistics to analyze biomedical phenomena is much later.

and regression analysis, and he discovered and explained fundamental statistical phenomena, among them the law of error and the regression toward the mean. We will return to these contributions in the next section, since are of paramount importance to explain the relationships between statistics and mathematics.

As commented on above, Galton was also responsible for the fruitful interest of biostatistics in explaining the apparent divergence between mendelian genetics and the evolution theory. This concern of Galton on the relationships between statistics, evolution and heredity was quite natural, since Galton was the cousin of Charles Darwin, author jointly with Alfred R. Wallace of the evolution theory. When Wallace (1855, 1858) and Darwin (1859) proposed their evolution theory, Mendel's work remained ignored and undiscovered, and the idea that an organism could pass on characteristics acquired during its lifetime to its offspring was the dominant heredity theory. However, the inheritance of acquired characteristics—due to Jean-Baptiste Lamarck [1744–1829] and also known as lamarckism—dissatisfied Galton, who began to apply statistics techniques to study continuous traits and population scale aspects of heredity on the basis of the natural selection hypothesis established in the evolution theory. On these aspects, Galton was methodologically influenced by Quételet—Galton himself fully recognized the previous contributions of Quételet and pursued the application of a bell-shaped distribution of characteristics identified by Quételet to the analysis of heredity—and by Wallace from the biological point of view, which contrary to Darwin strongly rejected the lamarckian idea of inheritance of acquired characteristics, something that Darwin had not ruled out. Galton's efforts to statistically demonstrate the mechanism of natural selection were continued by Karl Pearson [1857–1936] and W.F.R. Weldon [1860–1906], who persisted in working on the basis of the existence of continuous traits to explain the role of natural selection on heredity, and who on these premises founded the biometric school, of paramount relevance in biostatistics[4].

At this point, Gregor Mendel's ground-breaking work was rediscovered by Hugo de Vries and Carl Correns in 1900, providing arguments in principle observed as incompatible with natural selection and the continuous variation of characteristics observed by Galton, Pearson, Weldon and their biometric disciples. Indeed, the discoveries of the early geneticists were difficult to reconcile with the observed gradual and continuous evolution and with the mechanisms of natural selection, favoring saltationism and mutation by jumps instead. Mendelian evidence was indisputable, but so was the continuity of variation of organisms found by the biometric school, and the result was the coexistence over more than 20 years of two contradictory theories.

Biostatistics became crucial in solving this scientific dispute. The starting point was the work by the geneticist T.H. Morgan, which connected the mendelian genetics with the role of chromosomes in heredity, demonstrating that, rather than creating new species in a single step, changes in phenotype increased the genetic variation in

[4] Among other contributions, Weldon coined the term *biometry* and, jointly with Pearson, founded the highly influential journal *Biometrika*. Pearson is the creator of the *product-moment correlation coefficient* and *Pearson's chi-square test*.

a species population. On this basis, the biostatistician Ronald A. Fisher [1880–1962] elaborated a rigorous statistical model showing that, as a consequence of the action of many discrete genetic loci, the continuous variation of organisms observed by the biometric school, could be the result of the mechanisms of mendelian inheritance and natural selection. In a series of papers started in 1918 with the article *The Correlation between Relatives on the Supposition of Mendelian Inheritance* and culminating in 1930 with the book *The Genetical Theory of Natural Selection*, Fisher carried out an extremely acute statistical analysis of all these questions, which led to the reconciliation of Mendel genetics and the evolution theory and to the foundation of the Modern Evolutionary Synthesis, to a great extent the current paradigm in evolutionary theory. In addition, Fisher developed important statistical techniques, such as the analysis of variance, and is the father of the F distribution and several statistical tests.

At present, biostatistics strongly relies on the development of an evolutionary theory and on the work and methods of all these researchers we have quoted, but is much more than this. Indeed, biostatistics is playing a growing role in current medical and biological investigation, and today almost all medical or biological research papers use statistical methods and techniques. Through instruments such as descriptive measures, hypothesis tests, estimation, regression, and stochastic modeling, among others, biostatistics clarifies a large number of biomedical questions, helping to prove medical hypotheses, identify risk factors, recognize cause-effect relationships, discover explanatory variables, etc. As we have seen, biostatistics was born to satisfy the needs of biology and medicine, but its results in turn have contributed to the development of these sciences in which it was applied. As will be shown in the next sections, this has happened not only for the evolutionary theory, but also for the whole content of medicine and biology, in which that related to cancer will be given special consideration.

1.3 Biomathematics: Historical Notes

Unlike statistics, a science that from its origins joined biology and medicine in a natural and almost instantaneous way, mathematics was applied to the study of biomedical questions many centuries after its birth as science. Indeed, what are considered to be the first works in biomathematics, by the Italian mathematician V. Volterra and the US mathematician A.J. Lotka, were written in 1924 and 1926, that is several thousand years after the earliest uses of mathematics and almost 100 years after the opening work in biostatistics. Biomathematics is, undoubtedly, a very recent scientific discipline.

The motives for this delay in the application of mathematics to biology and medicine are multifarious. The first reason is the traditional consideration of biomedical phenomena as non-deterministic events, therefore non susceptible to mathematical deterministic description and only appropriately described by means of

statistical approaches. The second cause is the great complexity inherent to biomedical behaviors, characterized by a multiplicity of dynamic interrelationships between the involved entities difficult to mathematically formalize. And the third motive, related to the former, is the non existence prior to the second half of the 19th century of sufficient mathematical knowledge to tackle the mathematical formulation of biomedical phenomena.

Let us comment in more detail on these reasons for the relatively recent—and late—application of mathematics to biology and medicine. With respect to the first, the reluctance of biomedical scientists to use mathematics in the study of biomedical phenomena, it is worth noting again the different nature of statistics and mathematics. As explained in Sect. 1.1, statistics focus on the numerical and logical analysis of non totally predictable phenomena, while mathematics does so on the numerical and logical analysis of predictable behaviors. Given the attributes of biomedical events, patently non totally predictable, the general unwillingness and hesitation of the biomedical community to accept mathematics as a valuable and legitime analytical tool for analyzing biological and medical questions is understandable. On the contrary and as explained in Sect. 1.2, there was not opposition to the use of statistics, enthusiastically adopted from its origins as a valuable instrument to describe and interpret biomedical behaviors.

However, during the 19th and 20th centuries, as statistics was evolving and self-improving, stochastic behaviors in general, and biomedical conducts in particular, began to be interpreted as the mixed result of some deterministic regularities affected by a set of uncontrollable or unknown elements, these last of non-deterministic or stochastic nature. Regression analysis is a good illustration of this evolution. The term *regression* was coined by Galton (1877, 1907), who in his anthropometric studies detected a prevailing tendency—or *regression*—in the height of human individuals towards the mean value. This biological discovery of a regularity behind a stochastic behavior corroborated a general finding in statistics: As the mathematicians and astronomers Legendre (1805) and Gauss (1809) had previously documented in their studies of the planet orbits, movements in principe non totally predictable contain a well defined and deterministic central tendency. Applying the statistical methods and techniques proposed by Legendre (1805) and Gauss (1809), Galton started the application of regression analysis to study biomedical phenomena. The idea was the same as that in the work by the mathematicians Legendre (1805) and Gauss (1809): to find the deterministic component that underlies non totally predictable behaviors.

As a result, Galton (1877, 1907) opened up a new interpretation of the observed biomedical conducts: the biological and medical phenomena are of stochastic nature—and then they are non totally predictable—due to the existence of random elements, derived from measurement errors and unknown explanatory variables, that add to a well defined and deterministic central tendency. This interpretation not only justified the application of regression analysis in biology and medicine, a question that will be commented on in Sects. 3.9 and 4.2, but also conferred a role to mathematics in explaining biomedical behaviors: if in the biomedical phenomena there exists a deterministic central tendency, there is a component in the biological and medical conducts that can be mathematically described.

It is not then strange that simultaneously to the work by Galton (1877, 1907) applying regression analysis in biology and medicine and suggesting the presence of biomedical deterministic laws, mathematics began to be accepted as a valuable tool to explain biomedical behaviors. As we shall see, although mathematics was already present in medicine and biochemistry since the middle of the 19th century, the definitive impulse came from biology a few years after the work by Galton, in the 1920s. During this decade, the Italian biologist Umberto D'Ancona [1896–1964], by then researcher at the universities of Roma and Siena, observed that the captures of selachii—sharks, rays and skates—in italian seaports had unusually increased between 1914 and 1923, much more than the captures of their prey. D'Ancona argued that World War I had originated a decrease in the number of fish captures and then an increase in the population of fish preyed on by selachii. As a consequence of higher food resources, the population of selachii increased, as well as the number of selachii captures. However, an unanswered question remained: Why was a parallel increment in the captures of their prey not observed?

D'Ancona was the son-in-law of the Italian mathematician Vito Volterra [1860–1940], also a researcher at the university of Roma, and asked him for an answer to the problem. Volterra initiated his research on the subject by the end of 1925, and in 1926 found a response, published in two scientific papers: "Variazioni e fluttuazioni del numero d'individui in specie animali conviventi", and "Fluctuations in the abundance of a species considered mathematically".

Volterra's (1926a,b, 1931) mathematical studies on the relationships between prey and predator species crystalized in a model known as *Lotka-Volterra predator-prey* model, basically a system of differential equations. A similar system with the same mathematical properties had been previously proposed by the US mathematician A.J. Lotka in 1910 to describe some particular chemical reactions [Lotka (1910)], and later, to explain the behavior of specific organic systems [Lotka (1920)] and the interaction between prey and predators [Lotka (1925)]. This is why, jointly with V. Volterra, A.J. Lotka is considered co-father of the prey–predator model and co-founder of biomathematics. Indeed, Lotka's (1925) book "Elements of Physical Biology" is considered the first book on biomathematics.

The Lotka-Volterra model constitutes the origin of mathematical modeling in biology and medicine. As we have pointed out, it definitively broke the reluctance of biologists and medical scientists to accept the possibility of a mathematical approach for studying biomedical phenomena. This acceptance was neither easy nor immediate. Indeed, although after the work by Volterra (1926a,b) the interest on biomathematics increased among the scientific community, this first phase of attention was promptly followed by strained polemics about the legitimacy and properness of mathematical analysis in biology and medicine, and Volterra was even unable to publish in english his more important work on biomathematics, the book "Leçons sur la théorie mathématique de la lutte pour la vie".

However, the step had already been taken, and biomathematics began to be considered as the natural continuation of biostatistics. As the exiled Russian biomathematician V.A. Kostizin (1937) asserted[5]:

> Mathematics has entered into natural sciences through the door of statistics, but this phase must make way for the analytical phase as has happened in all the rational sciences. The role of statistical methods is to clear the field, to establish a certain number of empirical laws, to facilitate the step from the statistical to the analytical variables. This is an enormous and important task, but when it has been done, the word belongs to mathematical analysis, which, within this phase of formation of a rational science, is the only scientific field able to explain the causality behind the phenomena and to deduce from it all the logical consequences.

The Lotka-Volterra model is the first example of this attempt to provide the complex dynamic interrelationships that appear in biology and medicine with a deterministic and reductionistic mathematical approach. As we have commented, it is made up of a system of differential equations which, thanks to its dynamic properties, allows the problem of the relative increase in predators with respect to prey when human fishing activity decreases to be explained. The model would have been impossible to solve without the previous development during the 19th and 20th centuries of the mathematical theories of differential calculus and differential algebra. Indeed, only with the knowledge of these theories is it possible and feasible to tackle the mathematical formulation of biomedical behaviors. By their very nature, the dynamic and complicated interconnections between bioentities that characterize biomedical phenomena vary through time and space and depend on their current status, and only through a system of differential equations is it possible to capture and describe such interdependencies. Although the use of systems of differential equations goes back to the works by I. Newton [1643–1727], L. Euler [1707–1783], P. S. Laplace [1749–1827] and J.L. Lagrange [1736–1813], only after the investigations and the results provided by the mathematicians A.L. Cauchy in the 1820s, A. Cayley and C.G.J. Jacobi in the middle of the 19th century, and C.E. Picard and H. Poincaré at the end of the 19th and the beginning of the 20th centuries, was it possible to solve and analyze the peculiar systems of differential equations that describe biomedical phenomena. Due to these complexities, one inherent to the nature of the analyzed behaviors and the other to the mathematical knowledge, methods and techniques necessary to formalize them, the emergence of the first work in biomathematics at such a late date as 1926 is not strange.

In any case, once the reluctance of the biologists and medical researchers was overcome and the technical difficulties solved, biomathematics commenced to gain supporters as a useful, valid and fruitful scientific discipline. The work by A.J. Lotka (1910, 1920, 1925) and V. Volterra (1926a,b, 1931) was continued by Kermack and McKendrick (1927), who applied the same approach based on a system of differential equations to describe epidemiological phenomena, and was extended by Holling (1959a,b) and Murdoch (1977), who widened the range of dynamic interrelationships admitting an explanation in terms of differential equation systems.

[5] Translation from the French by the authors.

In addition, this approach to biomedical behaviors based on the utilization of systems of differential equations joined the mathematical results and analyses obtained in chemistry and biochemistry during the second half of the 19th and the first half of the 20th centuries. As is well known, the application of mathematics to chemistry dates from the 18th century and the origins of modern chemistry. Indeed, the works by R. Boyle [1627–1691], E. Mariotte [1620–1684], A.L. Lavoisier [1743–1794], J. Charles [1746–1823] and L.J. Gay-Lussac [1778–1850] constitute a perfect illustration of how mathematics contributed to elucidate fundamental aspects of chemical behaviors. The fruitful use of mathematics in chemistry continued during the 19th and 20th centuries, and, influenced by the results obtained by Galton (1877, 1907) in biostatistics, soon began to focus on relevant biological and medical questions. Among the researchers inquiring into the mathematical formulation of biochemical phenomena, the physicians A. Fick [1823–1901], L. Menten [1879–1960] and L. Michaelis [1875–1949] stand out.

In fact, although Lotka and Volterra are considered the fathers of biomathematics, the German physiologist Adolf Eugen Fick is without any doubt the predecessor of the mathematical approach to biology and medicine. As happened in biostatistics with the figure of A. Quételet, eclipsed by the work of Galton, biomathematics has a pioneer in A. E. Fick, overshadowed by the extremely innovative proposal of Lotka and Volterra. In particular, Fick is well known in biomathematics due to two paramount contributions: Fick's law and Fick's principle. Fick's law of diffusion is a quantitative law under the form of a single differential equation that describes the flow of particles from the area in which they are highly concentrated to the regions with lower concentrations. Fick (1855a,b) postulated his law to explain the diffusion in fluids through a membrane, that is to describe an osmosis process, and circumscribed his analysis to the chemistry and physics fields. However, although Fick did not look for a direct biomedical application of his law, Fick's works on diffusion were undoubtedly inspired by his medical knowledge and intuition, and today, equations based on Fick's law are profusely used in biology and medicine to mathematically model transport processes.

The interest of Fick in applying mathematics and quantitative sciences to medicine led him to devise a technique for measuring cardiac output, a technique that, mathematically expressed, is known as the Fick principle. The mathematical formula of the Fick principle, obtained in 1870 (Fick 1870), is jointly with Fick's law one of the first significant successful results in the application of mathematics to biology and medicine. These two contributions would have been enough to ensure Fick a prominent place in biomathematics as an outstanding pioneer, but his substantial achievements in quantitative and mathematical medicine are much more extensive. As Shapiro (1972) states, Fick is a clear exponent of the passion for incorporating mathematics into medicine and of the benefits that this incorporation brings:

> Adolf Fick, talented in mathematics and physics, gave to medicine and physiology the precision and methodology of the physical sciences. . . . Fick was unquestionably the Columbus of medical physics. His pioneering supplied the instruments and methods of physics which blessed physiology with precision. The plethysmograph, the pneumograph, the pendulum-myograph, the collodion membrane, the dynamometer, the myotonograph, the cosine lever,

> an improved thermopile and an improved aneroid manometer were all Fick's innovations. His formula relating deformation of the cornea to intraocular pressure (the Imbert-Fick law) refined applanation-tonometry for the diagnosis of glaucoma. The most accurate applanation-tonometer used today, the Goldmann instrument, is based on Fick's studies. These and other interests can be seen in his monograph, Medizinische Physik, published in 1856, when he was 27 years old. It was the first book of its kind, and went through 4 editions. . . . . Fick's text begins with molecular physics (the diffusion of gases and water, filtration, endosmosis), continues with mechanics (articulations, statics and dynamics of muscle), hydrodynamics of the circulation, sound recordings of circulatory events, the problem of animal heat and the conservation of energy, optics and color vision and closes with the measurement of bioelectric phenomena.

A. Fick was not the sole physician guessing the huge potential that the mathematical and quantitative approach to medicine and biology contains. The physicians L. Michaelis [1875–1949] and M.L. Menten [1874–1960] are other excellent examples of the fecund convergence of mathematics, medicine, biochemistry and biology. Influenced by the work of Fick—L. Michaelis was as Fick a German physician, and Michaelis and Menten developed their research in Berlin—, Menten and Michaelis (1913) carried out a mathematical analysis of the enzyme kinetics, obtaining a dynamic model that describes the enzymatic reaction rates through an equation, the Michaelis-Menten equation. This equation provided biologists, physicians and biochemists with a powerful mathematical tool to analyze enzymatic reactions, and quickly changed the study of biochemistry. Its importance is such that, today, the Michaelis-Menten equation is considered as the foundation of the kinetic analysis of chemical reactions, and is one of the key-stones of enzyme chemistry. As happened with Fick, this was not the unique relevant contribution to biomathematics of Michaelis and Menten. Indeed, the quantitative, mathematical and technical achievements of Michaelis and Menten are numerous.

For example, M. Menten's application of mathematics and physics to the study of biochemical phenomena led to significant improvements of electrophoretic techniques (in fact, she conducted the first electrophoretic separation of proteins), and to the basis of histochemistry (Menten is considered as the mother of this scientific field). Michaelis developed numberless quantitative, physical and mathematical analyses in several aspects of medicine and biochemistry. In this respect, Michaelis is known for his quantitative studies of the susceptibility of the various races of mice for cancer transplantation; for devising the method of the hydrogen electrode to quantify the hydrogen ion concentration; for elaborating a mathematical and quantitative theory of the dissociation of amphoteric electrolytes; for his calibration of the uni-colored pH indicators; and for extending and improving the theory and practice of potentiometric measurements.

The passion and desire of Fick, Menten and Michaelis to give biology, chemistry and medicine the precision of a mathematical approach stimulated and inspired many other scientists in these scientific fields. For instance, in 1925 the botanist G.E. Briggs and the biologist J.B. Haldane extended the mathematical analysis of the enzyme reactions proposed by Michaelis and Menten [Briggs and Haldane (1925)]; in 1913 and 1914 the biochemists D.D. Van Slyke and G.E. Cullen provided the basis of the gasometric procedures for measuring concentrations and a mathematical

formulation of the kinetics of urease action similar to that by Michaelis and Menten [Van Slyke and Cullen (1914)]; and, in 1934, the physical chemist H. Lineweaver and the biochemist D. Burk devised a powerful and useful formal method for analyzing the Michaelis-Menten equation [Lineweaver and Burk (1934)].

This stream of physicians, biochemists and biologists engaged in incorporating mathematics and quantitative sciences to biology and medicine and who, inspired by the results in biostatistics, developed their activity during the end of the 19th and the beginning of the 20th centuries, joined in the 1920s the mathematicians interested in applying mathematics to describe medical and biological phenomena. The efforts of Fick, Michaelis, Menten, Briggs, Haldane, Van Slyke, Cullen and some other physicians, biologists and biochemists who, from 1855 to the first years of the 20th century and ahead of their time, guessed the importance of mathematics in biology and medicine, found their reward when, after the work by Lotka (1910, 1920, 1925) and Volterra (1926a,b, 1931), all the scientific community understood the legitimacy and necessity of a mathematical approach to biology and medicine: Biomathematics had been born.

As we have commented, the merit of Lotka and Volterra was to show that mathematics can not only account for explaining the relationships between variables but also deal with the multiple complex dynamic interrelationships that characterize biomedical phenomena by using systems of differential equations. In fact, single equations and differential equations had already been used in biomedicine since the work by Fick (1855a,b) and Michaelis and Menten (1913), but the use of systems of differential equations was completely unprecedented as well as groundbreaking, since it opened up the possibility to explain the complicated dynamic interactions between a multiplicity of bioentities. This accomplishment definitively broke the reluctance of biologists and physicians to accept mathematical approaches in biology and medicine. As a result, once biomathematics were given carte blanche as a scientific field, physicians, biologists, biochemists and mathematicians quickly began to share ideas and knowledge, making the mathematical analysis of almost any biomedical question possible. Today, biomathematical models are applied to the study of cellular systems, cell cycles, organogenesis, tumorigenesis, enzymatic reactions, protein synthesis, therapies, species populations, genetics, immunology, organ functioning, pharmacokinetics, and a large et cetera of subjects virtually covering all biology and medicine.

Without any doubt, this development has been based on the utilization of systems of differential equations. Of course other mathematical approaches, methods and techniques coexist and are applied jointly with this kind of systems, but it is undeniable that the main corpus of the biomathematical results derives from applications of systems of differential equations[6]. As will be explained in Chaps. 6 and 7, this is due to the capability of differential equations systems to provide a full explanation

[6] This constitutes an additional argument to consider Lotka and Volterra as the fathers of biomathematics.

in biomedical terms of the analyzed phenomenon, that is, an explication of the totality of its salient features, including the nature of the interrelationships between the involved bioentities and of the dynamic evolution of each specific bioentity involved.

The modeling of biomedical phenomena through systems of differential equations constitutes then a clear advance with respect to their biostatistical description. In addition, it opens up the possibility of controlling the modeled biomedical behaviors, a very relevant question in biology and medicine. Indeed, since in a system of differential equations all the involved variables have linked dynamics, it is feasible to govern the whole phenomenon by controlling a reduced number of variables, the so called state variables. This is specifically the aspect analyzed by the *optimal control theory*, developed in the 1950s and 1960s by R. Bellman (1957) and L.S. Pontyagrin (1962), and with evident and obvious applications in biology and medicine. Indeed, as a result of optimal control theory, if the dynamic behavior of a biological phenomenon is accurately described by a system of differential equations, it becomes possible to govern the described behavior by manipulating some of the bioentities involved. If we think of these manipulated bioentities as the medicines or drugs administered in therapies, applying optimal control results is perfectly feasible to design the optimal therapy, that is, to find the quantities of drugs to be dispensed in order to produce the desired (optimal) dynamics of the biological system.

Today, the application of optimal control theory to design optimal therapies is very widespread, and without any doubt represents the core of current biomathematics. It is enough to have a look at the recent research literature on biomathematics to realize that the design of optimal treatments constitutes the subject of the majority of the papers. In addition, optimal control theory opens a new interpretation of the biological phenomena as self-governed events, a promising novel perspective with interesting repercussions in biology and medicine as shown by Gutiérrez et al. (2009). All these aspects concerning the implementation of optimal control theory in biosciences will be analyzed and discussed in detail in Chaps. 8 and 9. The reader interested in completing these historical notes on optimal control is also referred to those chapters.

Equations, systems of equations and optimal control are the main mathematical tools used today in biology and medicine, but not the only ones. In this respect, a relevant mathematical approach coexisting with the aforementioned mathematical instruments is *game theory*. Game theory is a branch of mathematics born in the 1940s after the work by J. Von Neumann and O. Morgenstein (1944), that mathematically describes strategic behaviors in situations of conflict and/or concordance of interests. In biology and medicine, game theory has mainly been applied to qualitatively explain the behaviors of individuals in making choices that depend on the choices of others. For the purposes of this book, these individuals can be bioentities in a wide sense (genes, cells, organs, ..), patients, or even public health offices. Then, and as will be shown in Chap. 10, game theory appears as a valuable and useful tool in biomathematics for researchers, practitioners and public health politicians, since it helps not only to analyze pure biological and medical questions but also to design optimal therapies and optimal public health policies.

Further Readings We can not describe in detail all the historical and methodological relevant questions concerning biostatistics and biomathematics, since it exceeds the scope of this book, which exclusively seeks to orientate and to guide cancer researchers in the use of these scientific disciplines. We remit the reader interested in going deeper into those topics to more specialized publications and studies. The following list provides with some useful references.

For the methodological aspects inherent to Science and its attributes of openness, universality, integrity and honesty, see Popper (1934, 1963, 1972), Lakatos (1976, 1978), Kuhn (1970), Macrina (1995) and Russo (2010).

For historic details on biostatistics, the interested reader can consult Galton (1909), Eknoyan (2008), Pearson (1906, 1914), Box (1978) and Heyde and Seneta (2001).

Good bibliographic sources on the history of biomathematics are the articles by Israel and Millán Gasca (1993), Millán Gasca (2009), Shapiro (1972), Michaelis et al. (1958), Bochner (1958) and Baird Hastings (1976).

Chapter 2
Descriptive Biostatistics

Abstract In this chapter we briefly describe the main methods and techniques in descriptive biostatistics, as well as their application to the analysis of biomedical questions. With special attention to the study of cancer, this chapter provides a general understanding of the nature and relevance of descriptive biostatistical methods in medicine and biology, explains the design behind a biostatistical descriptive analysis, and stresses the paramount importance of descriptive statistics as the initial stage of any biomedical investigation.

2.1 Descriptive Statistics: The Starting Point

As explained in the former chapter, statistics can be defined as the scientific study of the numerical data that emerge from non totally predictable phenomena. Starting from this definition, we will distinguish between the two alternatives that biostatistics (and in general statistics) offers to extract conclusions and information from a set of data. The first analyzes the data without assuming any underlying structure for such data, and is called *descriptive statistics*, while the second, *inferential statistics*, operates on the basis of a given structure for the observed data.

When nothing is previously assumed for the observed data, the only possible task is to describe such data. This is why the branch of statistics pursuing this objective is called *descriptive statistics*. The descriptive stage is for obvious reasons the first step in any applied research, and must lead to a manageable numerical summary of the data concisely but accurately describing them. Indeed, the ultimate goal of descriptive statistics is not to explain the data or the event behind the data, but, on the contrary, to make a future explanation and interpretation possible.

Usually, and especially in medicine and biology, the observations originated by a phenomenon are large in number, dispersed, variable and heterogeneous, aspects that prevent the researcher from directly comprehending it. To make the understanding of the analyzed phenomenon possible, it is first necessary to present, arrange, measure, classify, describe and summarize the obtained data. These are specifically the tasks carried out by descriptive statistics, a discipline that without any doubt constitutes the entry door to any biomedical scientific investigation.

P. J. Gutiérrez Diez et al., *The Evolution of the Use of Mathematics in Cancer Research,* 17
DOI 10.1007/978-1-4614-2397-3_2, © Springer Science+Business Media, LLC 2012

Looking at how to describe, simplify and arrange the data, descriptive (bio)statistics makes use of six main instruments:

1. Statistical tables.
2. Graphical representations.
3. Measures of central tendency.
4. Measures of dispersion.
5. Measures of form.
6. Measures of correlation.

When the data obtained from a biomedical phenomenon refer to a unique aspect or variable, only the five first instruments are susceptible to application; in addition, measures of correlation are also possible when the researcher collects data relative to several characteristics. In the following sections we will succinctly comment on these descriptive statistical tools. As is logical, our intention is not to give a condensed course on descriptive biostatistics, but to provide a general understanding of the nature and relevance of descriptive statistical methods in medicine and biology.

We will first define statistical unit, statistical population, and statistical variable. A *statistical unit* is each member—or each element, in statistical terms—of the considered set of entities under analysis. This set of studied entities is known as *statistical population*, a denomination inherited from demography, the first application field of statistics[1]. Finally, a *statistical variable* is each aspect or characteristic of the statistical unit that is considered for study. These statistical variables can be qualitative or quantitative, depending on whether their nature is countable or not.

2.2 Univariate Descriptive Statistics

We will begin our discussion of the main descriptive biostatistical methods and techniques by considering the univariate case. As previously remarked, before explaining a biomedical phenomenon it must be described and characterized, descriptive analysis being the first fundamental stage in any applied research. As a result, in biology and medicine, most research programs have a set of descriptive statistical analyses as starting point.

Let us suppose that, with the ultimate goal of explaining a biomedical phenomenon, we have considered a population and measured a particular characteristic for each member of this population. Let N be the number of individuals in the population, let C denote the analyzed characteristic, and let C_i, $i = 1, 2, \ldots, I$ be the different values for this characteristic. We will denote the number of individuals

[1] Indeed, traditionally, the origin of descriptive statistics dates back to the demographic work by John Graunt (1662).

Table 2.1 Univariate statistical table

Values of the characteristic	Absolute frequency	Relative frequency
C_1	n_1	$f_1 = \frac{n_1}{N}$
C_2	n_2	$f_2 = \frac{n_2}{N}$
...	...	...
C_i	n_i	$f_i = \frac{n_i}{N}$
...	...	...
C_I	n_I	$f_I = \frac{n_I}{N}$
Total	$\sum_{i=1}^{I} n_i = N$	$\sum_{i=1}^{I} f_i = 1$

presenting the value C_i as n_i. This number n_i is the *absolute frequency* of the observed value C_i, and the ratio

$$f_i = \frac{n_i}{N}$$

is known as the *relative frequency* of the observed value C_i. Note that f_i is the proportion on the total population N of individuals presenting the value C_i, and that

$$\sum_{i=1}^{I} n_i = N, \qquad \sum_{i=1}^{I} f_i = 1.$$

The observed values for the characteristic and their absolute and relative frequencies are usually arranged in tables. When the number of considered characteristics is one, the table is called a *univariate statistical table*. Table 2.1 is an example of a univariate statistical table.

Usually, the information contained in this table is presented graphically under the form of histograms and cumulative frequency curves. A *histogram* is a graphical representation of the absolute or relative frequencies for each value of the characteristic. A *cumulative frequency curve* is a plot of the number or percentage of individuals falling in or below each value of the characteristic. As is obvious, the histogram shows the relative presence or weight of each value of the characteristic in the population, whilst the cumulative frequency curve shows, with respect to each value of the analyzed characteristic, the population percentages displaying equal or lower values.

Statistical tables, histograms and cumulative frequency curves are just different ways to present the obtained data. In fact, none of these three descriptive statistical instruments imply modifications or manipulations of the data, which are simply collected and arranged. Together with these purely descriptive methods, there exist functions of the data that describe and summarize them and that are of paramount importance in descriptive biostatistics. For the univariate case we are examining,

these functions describing and summarizing the observed data are of three types: the so called *measures of central tendency*, the *measures of dispersion*, and the *measures of symmetry and form* or *moments*. They are also called *statistics*, since this term—statistics—refers not only to the science we are discussing, but is also applied to any function of the observed data. It this therefore in this latter sense—statistics as a function or modification of the obtained data—that the term statistics is used to design these measures of central tendency, dispersion and form.

The main statistics of location are the mean—which can be arithmetic, geometric or harmonic—, the median and the mode. Dispersion measures are given by four main statistics—range, standard deviation, coefficient of variation and percentiles—, and finally, symmetry and form are measured by variance, semivariance, skewness and kurtosis. We will not define or describe in detail these statistics since this is not the purpose of this book. However, two questions are worth noting. First, all these descriptive measures of the observed data set are a direct consequence of the frequency distribution—or histogram—of the data. As explained, the frequency distribution is simply an arrangement of the data showing the frequency of each value, and constitutes all the obtained information. The frequency distribution reports on the percentage weight that each observed value has on the total set of data, and makes judgements and predictions on the observations set possible. For instance, from the frequency distribution, the probability of observing a value or an interval of values in the data set can be exactly measured, and which value is the more likely to be observed in the data set can be predicted. This information is condensed by the aforementioned statistics, which provide a numerical summary of the frequency distribution. Second, among all these descriptive measures, researchers must choose the most appropriate for their purposes. For instance, if researchers want to compare the dispersion of two different populations, the coefficient of variation and not the standard deviation is the pertinent measure. In other cases, transformation of the original data can be convenient to eliminate asymmetry or to make variances independent of mean, and therefore the appropriate type of mean must be chosen. For these aspects and many others, we refer the reader to any of the excellent textbooks on biostatistics available today.

A very good example of the fundamental role that the descriptive stage plays in dealing with a biomedical question and of how descriptive biostatistics may help in extracting medical conclusions is Russo and Russo's (1987b) paper "Biological and Molecular Basis of Mammary Carcinogenesis". In this paper, the authors opened up a research avenue on breast cancer after concluding that malignant phenotypical changes in human breast epithelial cells are inversely related to the degree of glandular lobular differentiation and glandular development of the donor gland, and directly related to the proliferative activity of its epithelial cells. To arrive at these conclusions, basic for the subsequent analyses of how and when breast cancer initiates, the authors carried out an exhaustive descriptive statistical analysis of the relevant ratios and variables, that help them in a decisive way. For instance, to support one of their hypotheses, namely that the malignant phenotypical changes in breast epithelial cells are inversely related to the degree of glandular lobular differentiation and glandular development of the donor gland, the authors begin by depicting the histogram showing the decrease with aging in the number of terminal end buds for two kinds

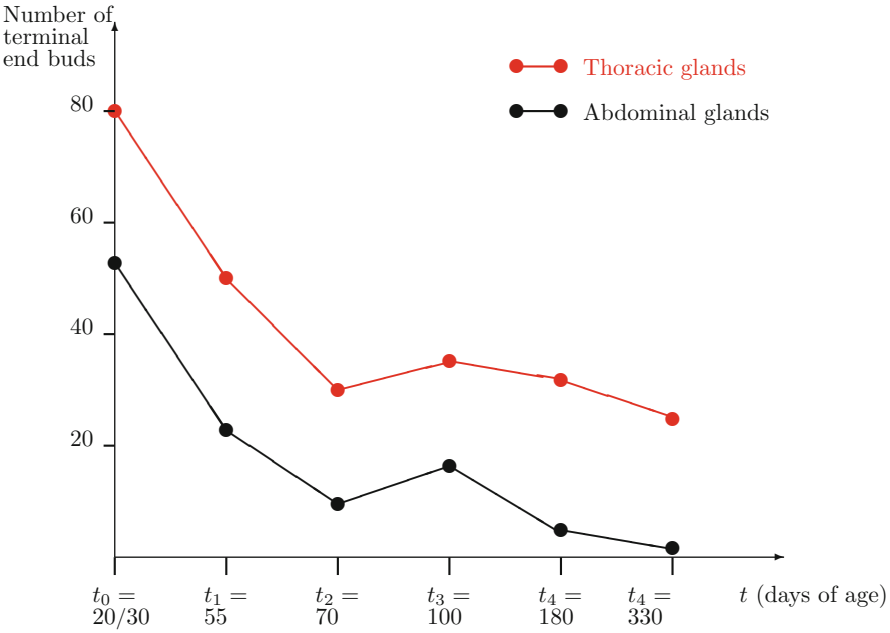

Fig. 2.1 Number of terminal end buds for thoracic and abdominal rat mammary glands. (Figure 25 in Russo and Russo (1987b))

of rat mammary glands, the thoracic glands and the abdominal glands. The evidence arising from this histogram is clear, since the number of terminal end buds for the thoracic glands always lies manifestly above the number of terminal end buds for the abdominal glands. Figure 2.1 reproduces the histogram in Russo and Russo (1987b) showing the number of terminal end buds for thoracic and abdominal rat mammary glands.

The following step to prove the hypothesis is to show that the thoracic glands are behind abdominal glands in development. This is again done through the appropriate histogram, which in this case represents the increment with aging in the number of alveolar buds and lobules, also for the two types of rat mammary glands. The conclusion from this histogram is also obvious, since the increments for the thoracic glands are always and evidently behind the increments for abdominal glands, as appears in Fig. 2.2.

Since (1) the terminal end buds are undifferentiated structures of the mammary gland, and (2) the increment in the number of alveolar buds and lobules is a direct sign of gland development, the final step to prove the hypothesis is to measure the tumor incidence for the two kinds of glands and to show that this incidence is greater in the thoracic glands than in the abdominal glands. This is also done by depicting a histogram showing the percentage of adenocarcinomas induced in the two types of glands, which allows the hypothesis to be proved: tumor incidence is greater in those glands located in the thoracic gland, which are less differentiated (with more terminal