ARBOR Ciencia, Pensamiento y Cultura
CLXXXIII 725 mayo-junio (2007) 353-361 ISSN: 0210-1963

STATISTICS IN BIOMEDICAL RESEARCH

Carmen Cadarso-Suárez
Biostatistics Unit, Department of Statistics and Operations Research, University of Santiago de Compostela, Spain

Wenceslao González-Manteiga
Department of Statistics and Operations Research, University of Santiago de Compostela, Spain

RESUMEN: *La Bioestadística es hoy en día una componente científica fundamental de la investigación en Biomedicina, salud pública y servicios de salud. Las áreas tradicionales y emergentes de aplicación incluyen ensayos clínicos, estudios observacionales, fisología, imágenes, y genómica. Este artículo repasa la situación actual de la Bioestadística, considerando los métodos estadísticos usados tradicionalmente en investigación biomédica, así como los recientes desarrollos de nuevos métodos, para dar respuesta a los nuevos problemas que surgen en Medicina. Obviamente, la aplicación fructífera de la estadística en investigación biomédica exige una formación adecuada de los bioestadísticos, formación que debería tener en cuenta las áreas emergentes en estadística, cubriendo al mismo tiempo los fundamentos de la teoría estadística y su metodología. Es importante, además, que los estudiantes de bioestadística reciban formación en otras disciplinas biomédicas relevantes, como epidemiología, ensayos clínicos, biología molecular, genética y neurociencia.*

PALABRAS CLAVE: *Estadística, Epidemiología, Ensayos clínicos, Bioinformática, Programas de formación.*

ABSTRACT: The discipline of biostatistics is nowadays a fundamental scientific component of biomedical, public health and health services research. Traditional and emerging areas of application include clinical trials research, observational studies, physiology, imaging, and genomics. The present article reviews the current situation of biostatistics, considering the statistical methods traditionally used in biomedical research, as well as the ongoing development of new methods in response to the new problems arising in medicine. Clearly, the successful application of statistics in biomedical research requires appropriate training of biostatisticians. This training should aim to give due consideration to emerging new areas of statistics, while at the same time retaining full coverage of the fundamentals of statistical theory and methodology. In addition, it is important that students of biostatistics receive formal training in relevant biomedical disciplines, such as epidemiology, clinical trials, molecular biology, genetics, and neuroscience.

KEY WORDS: Statistics, Epidemiology, Clinical Trials, Bionformatics, training programs.

1. Biostatistics as a discipline

There has been explosive growth in the development of statistical methodology over the past several decades. Research in medicine and public health has been both a beneficiary of this new methodology and a source of new problems, to the extent that statistics applied to medical research – biostatistics – can nowadays be considered a discipline in its own right.

In fact, biostatistics has become a defined branch of science that uses an intricate combination of statistics, probability, mathematics and computing to resolve problems in the biomedical sciences. Because research questions in biology and medicine are diverse, biostatistics has expanded its domain to include any quantitative, not just statistical, model that may be used to answer these questions. As a discipline designed to yield information, biostatistics may also be considered as a (highly-developed) branch of medical informatics, which in turn forms part of the developing field of bioinformatics. Consequently, biostatistics draws quantitative methods from fields including statistics, operations research, economics, and mathematics in general; and it is applied to research questions in fields such as public health (including epidemiology, nutrition, environmental health, and health services research), genomics and population genetics, medicine, and ecology.

The important role that biostatistics and biostatisticians play in the field of medical research has always been widely recognized by the biomedical community, and today statistics applied to medicine can be considered a successful model for the introduction of statistics into scientific practice. As an indication of its importance, the International Committee of Medical Journal Editors (ICMJE, http://www.icmje.org) is advised by biostatisticians of recognized prestige like DG

Altman, and includes in its "Uniform Requirements for Manuscripts Submitted to Biomedical Journals" a series of recommendations for the correct application and explanation of statistical methods used. Similarly, the biomedical journals of highest prestige –such as *The Lancet,* the *British Medical Journal,* the *Journal of the American Medical Association* and the *American Journal of Epidemiology* –include biostatisticians (notably specialists in epidemiology and clinical trials) among their associate editors and reviewers. In addition, journals of this type typically have a "Theory and Methods" section admitting articles about methodology, often statistical methodology. This situation is spreading to other biomedical journals.

Biostatistics has developed enormously in recent years, due to continuing advances in diverse biomedical areas and fields. For example, new problems in biomedical research have led to the development of new statistical methodologies that would not otherwise have arisen, and at the same time have favoured ingenious adaptations of classical statistical techniques to new contexts of application (DeMets *et al.,* 1994, and Zelen, 2006).

So biostatistics has over the years become a discipline in itself, enriching not only medicine but statistics in general. We only need to look at the current issues of prestigious statistical journals, such as the *Journal of the American Statistical Association,* or the *Journal of the Royal Statistical Society,* to see the influence that biostatistics has on new statistical methodology. In addition, specialized statistical journals like *Biostatistics, Biometrics, Biometrika, Statistics in Medicine* and *Statistical Methods in Medical Research* are considered highly prestigious within statistics, and occupy high positions in the Journal Citation Index ranking for the category "Statistics and Probability". In recent years, new journals have appeared in this ranking, such as the *Biometrical Journal,* the *Journal of Agricultural Biological and Environmental Statistics, Stochastic Environmental Research and Risk Assessment (SERRA),* and *Environmetrics,* covering applications of statistics in environmental sciences and public health. It seems likely that the number of journals in this ranking will increase in coming years. Finally, it is worth noting the publication by Wiley of the 8-volume *Encyclopedia of Biostatistics* (2005), whose editors are biostatisticians of high repute. This work extends and broadens the biostatistical content of Wiley's highly respected *Encyclopedia of Statistical Sciences.*

2. Why is statistics necessary in medicine?

As we have noted, the growth of quantitative methods in the biomedical sciences has made biostatistics a key component in many research areas. Let us here consider why two major fields of medical research, namely epidemiology and clinical trials, depend on statistics as a fundamental tool in the achievement of their goals.

2.1. Epidemiology

Epidemiology is the study of how often diseases occur, and why, in different groups of people. Epidemiological information is used to plan and evaluate strategies to prevent illness, and serves as a guide for the management of patients in whom disease has already developed. Like the clinical findings and pathology, the epidemiology of a disease is an integral part of its basic description. The subject has its own special techniques for data collection and interpretation.

The connection between biostatistics and epidemiology has always been close. The early epidemiologists were physicians interested in the way in which diseases occur in populations, their causes, and their relationships with different medical and non-medical factors. The problems tackled by these pioneers were not confined to the study of epidemics, but extended to the evaluation of therapies. Many were skilled in quantitative reasoning and were knowledgeable about the statistical methods of their day. Then, from the 1930s on, epidemiology began to turn its attention to the study of chronic diseases. It became impracticable to use the same prospective research strategies that had been so obviously appropriate in the study of infectious diseases. And here it was statisticians, primarily Cornfield and Mantel, who provided a rationale for approximately valid inference based on case-control data. Biostatisticians became more involved in elaborating on the conditions for valid inference, with concerns about bias due to possible confounding factors. Furthermore, they began exploring other issues related to epidemiologic research, such as models –called dose-response models– for evaluating the effects of possible risk factors for disease. These effects are quantified by measures of association such as the odds ratio or relative risk, i.e. probabilistic concepts that need to be estimated appropriately, according to the type of study (case-con-

trol, cross-sectional, or cohort) used for each particular research project.

The variety of statistical methods required in epidemiology is immense, and has led to the appearance of numerous books dealing with applications of statistics in epidemiological contexts. We should also note that the prestigious *Encyclopedia of Epidemiologic Methods (Series in Biostatistics),* published by Wiley in 2000, dedicates major sections to the most significant contributions of biostatistics to the ongoing advances in epidemiological research.

2.2. Clinical trials

Clinical trials are an essential part of the medical research process. Through clinical trials, scientific discovery can lead to better ways of preventing, detecting, and treating diseases and medical conditions. Clinical trials are studies performed with human subjects to test new drugs or combinations of drugs, new approaches to surgery or radiotherapy, or new procedures to improve disease diagnosis or patient quality of life. Most hospitals now take part in clinical trials, which are only begun after laboratory studies have indicated that the new treatment or procedure is apparently safe, and that it has the potential to work better than existing options.

Recent years have seen a major increase in the importance of statistics in the field of drug development (Anello, 1990). Statistics plays an essential part at all stages of the clinical trial, from planning, through conduct and interim analysis, to final analysis and reporting. The statistician will typically devise the randomization schedules, advise on sample size, specify criteria for measuring treatment differences, and analyse response rates. The statistician will generally also be the link with the Independent Data Monitoring Committee.

Several emerging and recurrent issues relating to the drug development process merit particular mention. Emerging issues include ongoing changes to the FDA (Food and Drug Administration, http://www.fda.gov) regulatory environment, increasing internationalization of drug development, advances in computer technology and visualization tools, and efforts to incorporate meta-analytical methodology. Recurrent issues include the continuing development of statistical methods for handling subgroups in the design and analysis of clinical trials; alternatives to the "intention-to-treat' analysis in the presence of non-compliance in randomized trials; methodologies to address the multiplicities resulting from a variety of sources, inherent in the drug development process; and methods to assure data integrity. These issues pose a continuing challenge to the international community of statisticians involved in drug development. Moreover, the involvement of statisticians with different perspectives continues to enrich the field and contributes to improvements in public health. The important methodological contributions being made by biostatisticians to clinical trials research has led to the recent creation of a specific journal, *Pharmaceutical Statistics:* having appeared only in 2002, it is already in the JCR ranking for Statistics and Probability.

3. Statistical areas of interest in medicine

According to Armitage (1995), the areas of statistics that have most influenced medical statistics in recent years have been generalized linear models (including multiple linear regression), survival analysis, categorical data analysis, spatial statistics, and Bayesian methods (in diagnostic, epidemiological and clinical trials contexts). Meta-analysis, as a tool for evidence-based medicine, has likewise attracted considerable attention in recent years.

In 1994, Altman and Goodman suggested that the following new statistical methods were going to play a key role in biomedical research over coming years: (i) bootstrap (and other computer-intensive methods); (ii) Gibbs sampler (and other Bayesian methods); (iii) generalized additive models; (iv) classification and regression trees (CART); (v) models for longitudinal data (general estimating equations); (vi) models for hierarchical data; and (vii) neural networks.

In 1997, Houwelingen likewise suggested that the future would be marked by new biomedical applications (in epidemiology, historical data on oncological patients and their families; in ecology, spatial data); by new philosophies (causal models instead of randomized clinical trials; prediction versus prognostic modelling); new models (graphical chain models, random effects models); new computational facilities (with an impact on the other aspects); new techniques (graphic techniques, exact meth-

ods, pseudo-likelihood); and new forms of collaboration (databases for meta-analysis, Internet software, Internet publications).

Many of the predictions made by Altman, Goodman and Houwelingen are already a reality, and many of the new statistical techniques they argue for have already been applied in studies published in prestigious biomedical journals. However, while these new methods are already being used in biomedical research, not all are being widely used. One possible explanation for this situation may be the lack of standard software implementing these new techniques, so that most biomedical researchers view their practical application as difficult. It is certainly clear that new techniques may take some while to find their way into medical research, although there is some evidence to suggest that the speed of transfer has picked up in recent decades.

4. Some recurrent and emerging issues in biostatistics

Modern biostatistics presents a number of challenges in terms of both the continuing development of classical techniques and the creation of new techniques to resolve new problems. In this section we focus first on the methodology of survival analysis, as an example of the advances being made in classical biostatistics. We then turn our attention to various emerging fields that merit further research by biostatisticians: specifically bioinformatics, spatial statistics, neural networks, and functional data analysis.

4.1. New issues in survival analysis

In many biomedical fields, time-to-event data must frequently be analysed. In oncology, for example, interest typically centres on the patient's time of survival following a surgical intervention. The analysis of this type of survival experiment is complicated by issues of censoring and truncation (Klein and Moeschberger, 2003). Censoring occurs when we do not fully observe the patient's survival, due to death unrelated to the cancer under study, or disappearance from the study for some reason. Truncation basically occurs when some patients can't be observed for some reasons related to the survival itself. A common example of this is in HIV/AIDS studies of the incubation period (i.e. time from infection to disease). In these studies, the follow-up starts when the HIV virus is detected and the moment of infection is retrospectively ascertained.

Since 1959, the Kaplan-Meier curve has been a well-known estimator of the survival function, and it is extensively used in epidemiological and clinical research. The classical proportional hazards model of Cox (1972) is also widely used whenever the goal is to study how covariates affect survival. A generalization of the survival process arises when survival is the ultimate outcome but intermediate states are identified. In this situation, a sequence of events is observed, leading to more than one observation per individual. Intermediate states might be based on categorical time-dependent covariates such as transplantation, clinical symptoms (e.g. bleeding episodes), or a complication in the course of the illness (e.g. metastasis), or alternatively on biological markers (like CD4 T-lymphocyte levels). A classical approach for the analysis of data of this type is the timedependent Cox regression model (TDCM). Advantages of Cox's regression model include its easy interpretability and its availability in the majority of statistical packages.

In the 1990s, so-called multi-state models (MSMs) became available: these offer a better understanding of the disease process, leading to a better knowledge of how the time-dependent covariate affects the evolution of the disease. These modern models have several advantages over Cox's regression model. They offer a better understanding of the disease process, providing the hazard for movement out of one state into another (i.e. transition intensities), as well as many other types of information, including the mean time of sojourn in each state, and survival rates for each state. Covariates on transition intensities can also explain differences in the course of the illness among subjects in the population. Notably, MSMs can reveal how different covariates affect different transitions, something that is not possible with other models like the TDCM. In fact, it is very unlikely that the risk of death in patients who have received different treatments will be the same. Furthermore, the prognostic factors associated with the risk of death may differ depending, e.g., on the treatment received.

A considerable literature is nowadays available on the analysis of MSMs: see for example the books of An-

dersen *et al.* (1993) and Hougaard (2000), and references therein. The benefits of using the multi-state approach can be readily understood from a recent issue of *Statistical Methods in Medical Research,* published in 2002, entirely devoted to these models. Despite its potential, however, MSM is not used by practitioners as frequently as other survival analysis techniques. This is attributable to a lack of awareness of the software available, and perhaps some lack of understanding of the advantages of MSM compared to the simple Cox model. In our opinion, MSM approaches are likely to make a major contribution to our understanding of complex survival processes, and it is thus important to try to promote these approaches in the biomedical community, for example by including them in university statistics courses, and by developing user-friendly software applications.

4.2. Statistical methods in bioinformatics

A very rapidly emerging influence on biostatistics is the ongoing revolution in molecular biology. Molecular biology is now evolving towards information science, and is energizing a dynamic new discipline of computational biology, sometimes referred to as bioinformatics. Bioinformatics merges recent advances in molecular biology and genetics with advanced statistics and computer science. The goal is increased understanding of the complex web of interactions linking the individual components of a living cell to the integrated behaviour of the entire organism. The availability of large molecular databases and the decoding of the human genome may allow a scientist to plan an experiment and immediately obtain the relevant data from the available databases. This is an area in which statistical scientists can make very important contributions. For example, the use of micro-array technology has created novel statistical problems that will motivate much new biostatistical research. In recognition of this major new direction in our field several biostatistics departments (mainly in the U.S.) have already been renamed as "Biostatistics and Bioinformatics" departments.

To date, various computational and statistical methods have been developed and applied in bioinformatics. Recently, new approaches based on support vector machines (Vapnik, 1995) have been suggested (Ewens and Grant, 2005). Certainly, bioinformatics is characterized by enormous amounts of data, and a heavy reliance on computing. Correspondingly, advances in statistical methods necessary for analysis are following closely behind advances in data generation methods.

Statistical competence is for example of vital importance in transforming the huge amounts of functional genomics data into usable knowledge, addressing extremely complex biological issues such as (a) the design of cDNA microarray experiments (how should the samples be distributed on arrays to get the most accurate results?); (b) the conversion of images in datasets (what is the optimal way to grid and segment images, to relate spot intensities to background intensities, and to normalize intensity data within and between arrays); (c) the identification of similar groups of samples and/or genes using statistical clustering methods; (d) the discrimination between groups of samples or assignment of new samples to one of a number of groups, using statistical methods for discrimination and classification; and (e) the identification of genes that are differentially expressed, based on hierarchical Bayesian inference, classification modelling or statistical tests.

As we can see, the statistical methods required by bioinformatics present many new and difficult problems for the research community. To date, the methods of analyses used have been mainly ad hoc, and remain at the development stage. It is clear that, to achieve optimal results in bioinformatics, close cooperation between researchers in statistics and researchers in genetics, biochemistry, medicine and biology is essential. A major forum for multidisciplinary publications is the journal *Bioinformatics,* which currently occupies the number-one position in the JCR category Statistics and Probability.

4.3. Spatial statistical methods in health studies

The analysis of the geographical distribution of the incidence of disease and its relationship to potential risk factors has an important role to play in various kinds of public health and epidemiological study. This general area is referred to as "geographical epidemiology", and four broad areas of statistical interest can be identified:

(a) *Disease mapping* aims to produce a map of the true underlying geographical distribution of disease incidence, given "noisy" observed data on disease rates. This may be useful in suggesting hypotheses

for further investigation, or as part of general health surveillance and the monitoring of health problems: for example, in helping to detect the outbreak of a possible epidemic, or in identifying significant trends in disease rates over time.

(b) *Ecological studies* are concerned with associations between the observed incidence of disease and potential risk factors, as measured on groups rather than individuals, the groups typically being defined by geographical area. Such studies are valuable in investigating the aetiology of disease, and may help to identify future lines of research, and possibly preventative measures.

(c) *Disease clustering studies* focus on identifying geographical areas with a significantly elevated risk of disease, or on assessing the evidence of elevated risk around putative sources of hazard. Uses include the targeting of follow-up studies to ascertain reasons for observed clustering in disease occurrence, or the investigation of control measures where the aetiology of observed clustering has been established.

(d) *Environmental assessment and monitoring* is concerned with ascertaining the spatial distribution of environmental factors relevant to health, and exposure to them, so as to establish necessary controls or take preventative action.

Given the breadth and importance of the concerns in geographical epidemiology, it is not surprising that there has been considerable interest in this area in recent years. Much of this interest has been in the development of relevant statistical methods and techniques, and there is no doubt that this particular vein of research has been, and continues to be, a fruitful source of interesting statistical problems, motivating successful methodological developments within the discipline. Several issues of major statistical journals have been devoted to spatial statistical methods in health applications (e.g. the *American Journal of Epidemiology,* the *Journal of the Royal Statistical Society,* and *Statistics in Medicine).* There has also been a considerable volume of papers in the field published separately, in key journals including the *American Journal of Epidemiology, Biometrics, Biometrika, Environmetrics,* the *Journal of the American Statistical Association,* and *Statistics in Medicine.* In addition, a number of significant recent texts have been devoted to this subject area (see Cressie, 1993; Lawson and Cressie, 2000; Lawson 2006).

Finally, there have been various special initiatives concerned with statistical methods in geographical epidemiology. In particular, a considerable amount of statistical work has been conducted under the auspices of other agencies with long-term interests in geographical and environmental health issues, including the US Centers for Disease Control and Prevention (CDC), Environmental Protection Agency (EPA), and National Research Centre for Statistics and the Environment (NRCSE), as well as the Pan American Health Organization (PAHO), the World Health Organization (WHO), and various European Community government agencies.

4.4. Neural networks in medicine

Neural networks (NN) approaches in medicine have attracted many researchers, and these approaches have been implemented in several biomedical applications (see Ohno-Machado, 1996), including diagnostic systems, biochemical analysis, image analysis, and drug development. Neural networks, which simulate the function of human neuron networks, have potentially useful implementations in many applications domains. Unlike human decision-makers, NNs are of course unaffected by factors like fatigue, working conditions and emotional state.

NNs have been applied in various areas of medicine: they are widely used in diagnostic systems, for the detection of cancer and heart problems, and for the analysis of diverse types of medical image (including tumour detection in ultrasonograms, classification of chest x-rays, tissue and vessel classification in magnetic resonance images, estimation of skeletal age from x-ray images, and assessment of brain maturation). NNs are used experimentally to model the human cardiovascular system: diagnosis can be achieved by building a model of the cardiovascular system of an individual and comparing it with the real-time physiological measurements taken from that patient. NNs are also used as tools in the development of drugs for treating cancer and AIDS.

Neural networks are increasingly being seen as an extension to general statistical methodology, to be given full consideration alongside classical and modern statistical methods. Reviews arguing this viewpoint include those of Ripley (1993, 1996) and Cheng and Titterington (1994), and certainly it is a viewpoint that is being widely ac-

cepted within the neural networks community. However, and in spite of the statistical contributions to date, much research work remains to be done in this field, and there are a number of interesting open problems. For example, it would be of great interest to see the incorporation of modern flexible nonparametric regression models like generalized additive models (Hastie and Tibshirani, 1990), or the use of bootstrap inference in this context, and in our opinion areas like this merit ongoing research.

4.5. Functional data analysis and medicine

In recent years, because of technological progress, many scientific fields in which applied statistics is involved are now measuring and recording continuous (i.e. functional) data. Notably, many modern apparatuses allow biomedical researchers to collect samples of functional data (mainly as curves, though also as images).

Since functional data is presented in curve form, it is natural to use the curve as the basic unit in functional data analysis. Functional data tend to involve a large number of repeated measurements per subject, and these measurements are usually recorded at the same (often equally spaced) time points for all subjects, and with the same high sampling rate. Functionals of these curves, such as derivatives and locations and values of extrema, are sometimes also of interest. This situation is very common is areas of basic medicine like endocrinology, for example in studies of hormone levels after different drug doses; or in neuroscience, for example in studies to estimate the firing rate of a population of neurons, in which the unit of study is the firing curve of each individual neuron. Another example is in the study of growth curves where more than one characteristic of growth is observed, e.g. height and lung function. Other examples include continuous monitoring of biological functions where the predictor is an exogenous or an endogenous variable varying over time.

Functional data analysis is a very recent field of research in statistics, and interesting contributions and applications may be found in the excellent monographs by Ramsay and Silverman (2002), Ramsay and Silverman (2005) and Ferraty and Vieu (2006), or a special issue of the journal *Statistica Sinica* appeared in 2004. The diversity of this field is highlighted by the wide scope of methodological problems involved, from numerous applied scientific disciplines. Three specific types of problem are of particular current interest, in view of their links with nonparametric statistics: factorial analysis of functional data, regression with functional variables, and curve classification.

The aims of functional data analysis are usually of an exploratory nature -to represent and display data in order to highlight interesting characteristics, perhaps as input for further analysis. However, there may be other aims, including estimation of individual curves from noisy data, characterization of homogeneity and patterns of variability among curves, and assessment of the relationships of shapes of curves to covariates.

In spite of the important recent contributions in functional data analysis, the development of new tools to deal with functional data represents a significant challenge for the statistical community. Apart from the books already cited, a good starting point for thinking about new lines of statistical research and possible applications may be the special issue on functional data due to appear soon (2007) in the journal *Computational Statistics and Data Analysis.*

5. Training biostatisticians

As can be deduced from what we have said so far, the successful application of statistics in the biomedical sciences requires professionals with a high-level mathematical training, and at the same time a good understanding of the central issues of the biomedical sciences, including medical ethics, and of the systemic constraints of the research process. So a biostatistics professional needs to be trained not only in traditional statistical theory and methods (including special topics such as epidemiology, clinical trials and survival analysis), but also in bioinformatics and basic biology, as well as in communication and leadership skills. The demand on leadership is likely to be even greater than in the past, given the central role that biostatistics and bioinformatics now play in biomedical research.

Thus, a training programme in biostatistics must by necessity be interdisciplinary, connecting statistics training *per se* to an understanding of the basics of biomedical research. Trainee biostatisticians need to learn basic concepts of epidemiology, clinical trials research, genetics

and molecular biology, in order to be able to understand biomedical research problems and to make meaningful contributions to them. A modern training programme of this type, with an emphasis on bioinformatics, statistical genetics and computational biology, would profit from trainees spending time in biomedical laboratories to gain first-hand experience and insight into the nature of the real problems faced by these researchers. A major goal must be to train students to become independent researchers, advancing the field of statistical research and its application to biomedical research, both basic and clinical; and at the same time to be team workers intimately involved in the design and data analysis of collaborative biomedical projects.

Since many biomedical problems will likely require the development of new statistical methods, students should be capable of critically reading the theoretical and methodological literature in statistics, and of developing new methods as appropriate for the problems that they encounter. Students without such rigorous training are unlikely to be able to develop new methods or fully understand the strengths and limitations of those methods developed by others. The traditional component of biomedical courses will probably focus on areas of mathematical statistics including probability theory, inference, resampling methods (e.g. bootstrap), linear regression, analysis of variance, generalized linear models, survival analysis (including multi-state models), nonparametric methods, and data analysis. In addition, new methodologies like spatial statistics, neural networks, and smoothing regression methods (such as generalized additive models) are strongly recommended.

As we have noted, in order that biostatisticians can offer *new* statistical methodologies useful for the biomedical research community (in other words, methods that are not already incorporated into standard statistical software), they need to be competent at software creation. It is thus of interest for training programmes to include components dealing with relevant programming languages, like R, which facilitate the implementation of new methods (as well as the adaptation of existing methods to new biomedical contexts), and which thus help biostatisticians to create user-friendly software that can be readily used by researchers without a specialist statistical training. This is turn will favour the more rapid transfer of new statistical methods into the mainstream of biomedical research.

Also, since the goal is to train students not only to be excellent members of research teams, but ultimately to be independent biomedical researchers in their own right, students must gain some fundamental knowledge of a biological specialty area. This knowledge can be gained through formal course work or hands-on laboratory experience, or indeed by a combination of the two. For example, courses for a student focusing on statistical genetics might include statistical genetics, human genetics, and population genetics, as well as courses on areas of computer science like database structures, and bioinformatics.

Concern for the current situation of biostatistics in Spain led in 1999 to the organization of a Round Table on Statistics in Medicine, as part of IVth Congress of the Galician Society for the Promotion of Statistics and Operations Research. Under the direction of Professor Guadalupe Gómez, this round table constituted an interesting reflection on various questions of relevance to the profession of biostatistics, summarized in a valuable article in the journal *Questióo* (Abraira *et al.,* 2001). Specific issues discussed included the need to promote postgraduate courses in biostatistics, aimed at training biostatisticians of the highest rank within Spain. In line with this need, the Department of Statistics and Operations Research of the University of Santiago de Compostela launched in 2005 a new Master in Biostatistics (http://eio.usc.es/pub/master_bio), with the aim of training students in modern statistical methods for biomedical research, and at the same time giving them a solid grounding in the major areas of biomedical research, notably epidemiology, clinical trials, genetics and bioinformatics.

ACKNOWLEDGEMENTS

The work of Carmen Cadarso-Suárez and Wenceslao González-Manteiga was partially supported by grants MTM2005-00818 and MTM2005-00820 from the Spanish Ministry of Science and Technology (FEDER support included).

REFERENCES

Abraira, V., Cadarso, C., Gómez, G., Martín, A. and Pita, S. (2001): "Mesa Redonda: La Estadística en la Investigación Médica", *Qüestiió,* 25, 121-156.

Altman, D. G. and Goodman, S. (1994): "Transfer of technology from statistical journals to the biomedical literature: past trends and future predictions", *Journal of the American Medical Association,* 272-129.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993): *Statistical Models Based on Counting Processes,* Springer, New York.

Anello, Ch. (1999): "Emerging and recurrent issues in drug development", *Statistics in Medicine,* 18, 2301-2309.

Cheng. B. and Titterington, D. M. (1994): "Neural networks: a review from a statistical perspective", *Statistical Science,* 9, 2-54.

Cressie, N. (1993): *Statistics for Spatial Data,* Revised Edition, Wiley, New York.

DeMets, D. L., Stormo, G., Boehnke, M., Louis, T. A., Taylor, J. and Dixon, D. (2006): "Training of the next generation of biostatisticians", *Statistics in Medicine,* 25, 3415-3429.

Ewens, W. J. and Grant, G. (2005): *Statistical Methods in Bioinformatics (Statistics for Biology and Health),* Springer, New York.

Ferraty, F. and Vieu, Ph. (2006): *Nonparametric functional data analysis: Theory and practice,* Springer Series in Statistics, Springer, New York.

Greenhouse, S. W. (2003): "The growth and future of biostatistics: (A view from the 1980s)", *Statistics in Medicine,* 22, 3323-3335.

Hastie, T. J. and Tibshirani, R. J. (1990): *Generalized Additive Models,* London, Chapman and Hall.

Hougaard, P. (2000): *Analysis of Multivariate Survival Data,* Springer, New York.

Houwelingen (van), H. C. (1997): "The future of Biostatistics: expecting the unexpected", *Statistics in Medicine,* 16, 2773-2784.

Klein, J. P. and Moeschberger, M. L. (2003): *Survival Analysis: Techniques for Censored and Truncated Data,* Second Edition, Wiley, New York.

Lawson, A. (2006): *Statistical Methods in Spatial Epidemiology,* 2nd edition, Wiley.

Lawson, A. and Cressie, N. (2000): "Spatial statistical methods for environmental epidemiology", in *Handbook of Statistics,* Vol. 18, P.K. Sen & C.R. Rao (Eds.), Elsevier, Amsterdam, pp. 357-396.

Ohno-Machado, L. (1996): *Medical Applications of Artificial Neural Networks: Connectionist Model of Survival,* Ph.D Dissertation, Stanford University.

Ramsay, J. and Silverman, B. (2002): *Applied functional data analysis; methods and case studies,* Springer, New York.

Ramsay, J. and Silverman, B. (2005): *Functional data analysis,* 2nd edition, Springer, New York.

Ripley, B. D. (1993): "Statistical aspects of neural networks", in Barndorff-Nielsen, O. E., Jensen, J. L., Kendall, W. S. (eds.), *Networks and Chaos-Statistical and Probabilistic Aspects,* London, Chapman & Hall, 40-123.

Ripley, B. D. (1996): *Pattern Recognition and Neural Networks,* Cambridge University Press.

Vapnik, V. (1995): *The Nature of Statistical Learning Theory,* Springer-Verlag.

Zelen, M. (2006): "Biostatisticians, Biostatistical Science and the Future", *Statistics in Medicine,* 25: 3409-3414.

Recibido: 15 de diciembre de 2006
Aceptado: 20 de diciembre de 2006