STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:3409–3414
Published online 23 August 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2658

Paper Celebrating the 25th Anniversary of *Statistics in Medicine*

Biostatisticians, Biostatistical Science and the Future

Marvin Zelen*, †

Department of Biostatistics, Harvard School of Public Health, Harvard University, Boston, U.S.A.

SUMMARY

Biostatistical Science is in a 'golden period'. The definition of Biostatistical Science is the application of statistics, probability, mathematics and computing to advance our understanding of the subject matter in the biomedical sciences. Our field is experiencing unparallel developments due of the advances in communication and computing. We are becoming more global, we have resources which can expand educational opportunities for distant learning; the growth of quantitative methods in the biomedical sciences has made biostatistical science a key component in many research areas. What about the future? Are we receptive to change as many new scientific areas expand? Will the interaction between academia and industry likely to grow—especially in the training of future practitioners of biostatistical science. This paper discusses some of challenges facing our profession if we are to continue to be relevant in the biomedical sciences. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: biostatistical science; computational biology; distance learning; software

I would like to thank the Editors for inviting me to submit this paper to celebrate the 25th anniversary of the publication of Statistics in Medicine. Parts of this paper were previously presented before, in meetings of the International Chinese Statistical Association, the Eastern Mediterranean Biometric Society and a meeting held in Taejon, South Korea in conjunction with the 53rd meeting of the International Statistics Institute. This talk will discuss several emerging issues that affect the current and future practice of our profession.

I first attended graduate school in 1949. Rather than looking backward I thought I would look ahead and devote part of my paper to the future of Biostatistical Science.

This task should only be undertaken by those possessing a blend of courage and foolhardiness. To illustrate my point, it is only necessary to go back a hundred years and attempt to predict the future of Biostatistical Science in the year 1900. Of course, at that time, our specialty and even the term had not yet been invented. Perhaps one might question whether our field will exist in another hundred years. Biostatistics may evolve into other specialties and disciplines.

*Correspondence to: Marvin Zelen, Department of Biostatistics, Harvard School of Public Health, Harvard University, 655 Huntington Avenue, Boston, U.S.A.
†E-mail: zelen@jimmy.harvard.edu

Received 14 June 2006
Accepted 14 June 2006

I prefer to use the term Statistical Science to describe the practice of Statistics. By Statistical Science, I mean the application of statistics, probability, mathematics and computing to advance our understanding of a subject matter field. I refer to the practitioners of Statistical Science as Statistical Scientists, not Statisticians. The terms 'statistics' and 'statisticians' were popularized when statistics was concerned with 'political arithmetic'. Political arithmetic dealt with the study of records for the purpose of administration by government. This is very different from how Statistical Science is practiced today. It is noteworthy that many university departments of statistics have been renamed as departments of Statistical Science. The oldest such department located in University College, London was renamed as the Department of Statistical Science many years ago.

When the main field of application is in the biomedical sciences or agronomy we may often describe this activity as Biostatistical Science and its practitioners as Biostatistical Scientists. A large part of the activity of current-day biostatistical scientists deals with the observation, identification, data description and theoretical explanation of phenomenon in the life sciences. The endeavours are broader than methodology. The term 'Biostatistical Science' seems most appropriate to describe these activities. The vitality of our profession is dependent on the scientific component of our efforts. Without the emphasis on advancing knowledge, our profession will assume the role of a guild, which has as its primary mission the mutual aid and protection of its members in the marketplace.

Nearly all of us have ready access to enormous communication and computational capabilities which were undreamed of a few decades ago. Many of us are on the Internet every day. This has changed the way we practice our science. No longer do we tend to have collaborators who are mainly located in our same geographical region. Our close 'colleagues' are not necessarily in the adjoining offices, but tend to share the same interest as ourselves. We may be in close communication even though they may be located far away.

We have not fully taken advantage of the potential of our communication resources to educate our future biostatistical scientists. I propose that our profession assemble courses on the Internet that would be freely available. Many faculties have favourite courses which would be useful to wider audiences. It is routine in many universities for faculty to teach courses using slides. These slides are often in PowerPoint, slitex or in pdf formats and can easily be put on the Internet. Consequently, with guidance from a faculty member (or a knowledgeable person), a biostatistics student could have access to the same courses taught in major centres of biostatistics. The courses could range from the most elementary to the most advanced. Not only course materials can become available, but seminars can also be made available as well. The means to view slides in Corfu, Seoul or London for a seminar which was given in Boston last week are now readily available. The technology has advanced so rapidly that it is also possible to see a video of the speaker presenting the class or seminar. Many universities are video taping courses. These could be organized by our profession and made freely available worldwide. A model for making courses/seminars available on the Internet already exists in the field of epidemiology (http://www.pitt.edu/∼super1). At the University of Pittsburgh, Professor Ron LaPorte and his colleagues have collected more than 2000 courses and seminars, which are freely available on the Internet. It is a worldwide effort with courses being contributed from many different parts of the world. I recently contributed a course entitled 'Stochastic Processes in Public Health'. It was put in proper format by someone in Russia. The courses are mainly in English, but there are several courses in Chinese, French, Portuguese, Russian and Spanish. A few lectures are available in Arabic, Croatian, Georgian, Indonesian, Macedonian, Serbian, Korean and Japanese. The contributors come from 151 countries. Our profession should follow the model of the epidemiologists and make 'distance learning' available to all.

DOI: 10.1002/sim

Biostatistical Science is enjoying unparalleled developments. The need and demand has never been greater—especially in the U.S. The position paper by DeMets *et al.* [1] describes the growing shortage of biostatistical scientists in the U.S. This shortage will become worse with time as the demand for biostatisticians continues to increase and the availability of training funds for graduate training in biostatistics continues to diminish. A 'bright' side to this 'unhealthy' situation is that salaries will continue to escalate. We need only look at any of the current issues of statistical journals to see the influence that the practice of Biostatistical Science has on new statistical methodology. In my view, Biostatistical Science is at the cutting edge of many new developments in statistics. This research is mainly motivated by problems in the health sciences. It reminds us that statistical theory cannot be separated from the practice of statistics. When they are separate, the theory is likely to be of little consequence and the practice runs the risk of being unsound. Some of our colleagues describe themselves as mathematical statisticians or applied statisticians. The former does not encounter data and the latter does not relate to theory. Such terms are outmoded and should be discarded.

This new century has been described as the 'Information Century'. The widespread availability of computing has led to the creation of many new databases. However, there is a distinction between data and information. Converting data into information requires insight, skill and training in the theory and practice of statistics. Our every day lives are being affected by predictions of weather, trends in the stock market, results of a national census, reports on new therapies for diseases, etc. I believe the mark of an educated person in this new century will be in the ability to reason with numbers. In nearly all medical schools, students are required to satisfy a biostatistics requirement. Most of the top medical journals have statistical editors in recognition that a great deal of modern medical research is quantitative. Our courses in biostatistics must meet the challenge of helping students gain skills in quantitative reasoning not only in science, but also in every day life.

Many countries are greatly influenced by activities and events in the U.S. The broad reliance of the health sciences on quantitative methods in the U.S. is influencing similar activities in other countries. Our profession is becoming very much a 'global profession'. We are witnessing hospitals located in different countries entering patients on common clinical trials. Beginning in 1991 there have been conferences every two years by representatives from Europe, Japan and the U.S. These conferences are called the International Conferences on Harmonization and are commonly referred to as the ICH. Attendees mainly come from the industry and government regulatory agencies. It is noteworthy that few academics are involved with the ICH. The goal is to draft guidelines to make global drug development more efficient. The ICH has already adopted a document entitled 'Statistical Principles for Clinical Trials' [2]. The pharmaceutical industry is clearly a global industry and the need for biostatistical scientists, as now felt in the U.S., will also be felt in other countries. Many multi-national pharmaceutical companies have biostatistical staffs—in some instances these are quite large. Augmenting the growth in the industry, many countries have created agencies to monitor the approval of new drugs. Many are modelled on the equivalent of the U.S. Food and Drug Administration. Unfortunately, the conflicting guidelines between countries have hampered the initiation of clinical trials. One of the goals of the ICH is to attempt to have the government guidelines somewhat uniform so that inter-country clinical trials may be carried out. These developments will not only create more job opportunities, but will also make our profession more important in assessing the safety and efficacy of new drugs.

This growth may be especially important in the Pacific Rim countries. Japan is the second largest consumer of drugs in the world. China represents a huge potential market. Yet there is

DOI: 10.1002/sim

relatively little clinical trial activity in these countries. It is not clear whether the results of trials carried out in the West directly apply to the inhabitants in the Far East. There exist opportunities to include the Pacific Rim countries in the planning of trials to evaluate new therapies, not only as a source of patients, but also to determine whether drug therapies are dependent on ethnic groups. Furthermore, some aspects of 'Traditional Medicine', as practiced in the Far East, may be beneficial and can augment western medicine. In 2004, I attended a conference in China where the main topic was the design of clinical trials to generate data to determine the benefits of traditional Chinese medicine. It is clear that there may be significant growth in clinical trial activity in the Far East generating scientific evidence on the safety and benefits of some of the practices of 'Traditional Medicine'.

The U.S. Food and Drug Administration (FDA) has exerted great influence on the employment of biostatistical scientists in the U.S. This demand is high and is unlikely to be abated in the near future. This need has provided opportunities for international students to seek biostatistics graduate training in the U.S. For example, in my own Department we have currently enrolled students from: Argentina, China, Greece, Mexico, Portugal, Spain, Taiwan, Korea and Singapore. After completing their studies many of the students remain in the U.S. However, as employment opportunities for biostatistical scientists grow in other countries, I would expect that an increasing number of international students would return to their native countries and exert their influence on biostatistical training and practice in their own countries. I can foresee that some of the leadership in Biostatistical Science, which is in the U.S., the U.K. and Europe, will gradually be shared by countries in the Far East. New academic programme in biostatistics are likely to be initiated in many Universities—especially in the Pacific Rim countries. This direction is likely to be greatly accelerated if our profession is successful in utilizing the Internet for making courses and seminars available on-line.

Another theme that is emerging very rapidly and is affecting our discipline is the revolution in molecular biology. Molecular biology is now evolving into Information Science and has energized an emerging new discipline of Computational Biology, sometimes referred to as Bioinformatics. This discipline is concerned with the discovery and implementation of algorithms that facilitate the understanding of biological processes. The principal tools are computing, statistics and so-called 'machine learning' techniques. The field is characterized by heavy reliance on computing and generating enormous amounts of data. The methods of analyses are mainly ad-hoc and are still in a 'development' stage. Not too long ago, one would have a hundred bench scientists generating new data and only a handful of individuals analysing the data. We are now in the reverse situation. A small group of bench scientists can generate an enormous amount of data that may require large numbers of data analysts. 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 activity in which statistical scientists can excel. 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 have already been renamed as Biostatistics and Bioinformatics. Looking far ahead to the future, the Biostatistical Scientist would look back at our time and comment on the 'quaint' way we planned and analysed studies. They would be intrigued why it was necessary to carry out clinical trials enlisting hundreds, if not thousands of patients, taking many years to complete, when in fact they need only look at a therapeutic drug's chemical composition and predict benefit, conditional on knowing an individual's genome profile. It sounds like science fiction, but going back a hundred years, could anybody predict current activities in our profession?

DOI: 10.1002/sim

I believe many biostatistical scientists will identify themselves with application areas. They will regard themselves as primarily bioinformatic scientists or clinical trials scientists and only secondarily with biostatistics. The biostatistics profession continues to be in a golden period, mainly because we are major contributors to the advancement of the health sciences. A major issue for the future is how we can continue to be relevant as the needs of the health sciences change. Biostatistics is driven by being an applications-oriented field. As needs change so must our training. I can foresee that future needs in biostatistics will require significant training in bioinformatics, biology and modern genetics. There will be less emphasis on traditional mathematical statistics and much more attention to data analysis—especially dealing with large databases. However, universities are relatively conservative institutions and are slow to change. Science and technology are moving at a very fast pace. Universities must be able to face the challenge of changing curricula to take advantage and keep pace with new scientific opportunities. Otherwise, we are likely to find our profession less relevant in the future.

In 1982, at a meeting of the International Biometrics Society, I was invited to participate in a programme entitled 'The Future of Biostatistics'. My address was published the next year in Biometrics with discussion, Zelen [3]. In my talk I stated, 'The future of Biostatistical Science will be intimately related to computing'. I went on to cite my reasons and strongly recommended that a significant amount of training be devoted to computing. Two of the five discussants (the late Professors B. Greenberg, Univ., N.C. and S. Greenhouse, George Washington University, disagreed with my view. Professor Chin Long Chiang, University of California (Berkeley), later published a paper vigorously arguing that my remarks 'overemphasized the role of computing' [4].

It is now 25 years later! I still hold the same view on the role of computing—not only in Biostatistical Science, but also in Statistical Science. At that time, I had described the history of development of statistical software in four stages. The final stage, referred to as 'stage IV' described automatic data analysis systems. By this term I meant that the user will input a set of 'stylized' questions dealing with various hypotheses or models. The system will automatically choose one or more appropriate statistical techniques and give the answers to the stylized questions. Also, the computer will indicate various caveats or cautions relating to possible shortcomings of the methodology—for example, concerning assumptions, robustness or approximations—and how they might affect the conclusions. I had predicted this stage would arrive in about a decade. This evolution in computer aided automatic data analysis has yet to manifest itself. However, we are seeing the beginnings of such automatic data analyses in the monitoring of clinical trials. In some of these trials elementary interim analyses are being carried out automatically.

One of the problems we are experiencing in computing is the increasing cost of software. In former times we were accustomed to have access to whatever software was needed as cost was a relatively negligible item. This is no longer true today. I would hope that in the future there is a greater trend to make available open software. An ideal example is R. Robert Gentleman (Fred Hutchison Cancer Center) and colleagues from all over the world are owed a vote of thanks by our profession for their insight and ambition to launch R. In contrast, I know of one software system for clinical trials that charges over $50 000 with costly annual maintenance fees. Routine general-purpose statistical software may cost in the neighbourhood of $1000. I believe this system may change with more open software being made available. However, there may be some specialized software that will not be freely available. One way of having access to such specialized statistical software is to have it available on the Internet. The user will not have the software resident on the user's computer. However, with a password, he or she will be able to use the software and payment will be made by amount of use. It is expensive to write good software and software

DOI: 10.1002/sim

developers should be fairly compensated. However, a system has to evolve that does not prevent use of the software because of costs. A corollary to this problem is that students should have access to the most advanced software. It is incumbent on software developers to provide versions of their software, which would be suitable for classroom use. In the long run, software developers have much to gain by having students familiar with their software.

I wish to comment on relations between the University and the Industry. Leadership in biostatistics is mainly centred in universities. However, the industry is hiring many young and talented biostatisticians. As they become more senior we are likely to witness increasing leadership in the future directions of Biostatistical Science from those employed by the industry. This will undoubtedly result in closer relationships between the industry and the universities. We, in academia, should take steps to hasten this process. Both universities and industry share many goals. It is inevitable that the industry will begin to participate in the educational process. Parallel to the increasing influence of the industry in the educational process is the need for the industry to recognize that there is an obligation to support academia by providing resources for student fellowships and academic research. In the long run, this will enormously benefit the industry.

Finally, I wish to remark on the role of Statistical Science on policy issues. We have much to contribute to policy. I urge our profession to regard this as a major goal for the future. Many policy decisions are based on quantitative information. It is timely for our profession to expand our role, by not only being responsible for the collection and interpretation of data, but also to take a greater leadership role in helping make policy. Policy issues in health care are becoming dominant in both the internal and external affairs of many countries. It is only necessary to refer to the worldwide AIDS epidemic, the failure of the U.S. to endorse the Kyoto treaty, the outbreak of Sars in the Far East, the concern with the spread of the Avarian virus, and the recent outbreak of polio in some of the Arab countries. It is necessary to develop coherent and evidence-based policies to deal with these problems—The inputs to the decision makers are usually data, perhaps refined and integrated into quantitative models. Such models may predict the speed of the spread of an epidemic, the sub-populations which are most vulnerable, the magnitude of the demand for vaccines, etc.... I would hope that in the future members of our profession will take a more active role in helping shape health policy—especially at the international level. However, we must be cautious and sensitive to the possibility of politics tempering our conclusions.

I congratulate those past and current editors who played an important role in making 'Statistics in Medicine' an important journal for our profession. The success of the journal is a reflection of the vitality of our profession. Perhaps the next 25 years will be even more successful and those future editors would once again extend an invitation to me to contribute an article about the future.

ACKNOWLEDGEMENTS

I would like to thank Professors D. DeMets, S. L. Lagakos and L. J. Wei for making comments and suggestions on a preliminary draft of this paper.

REFERENCES

1. DeMets D, Louis T, Lagakos S, Dixon D. Increasing support for training of the next generation of biostatisticians: a call to action. *Statistics in Medicine*, in press.
2. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use. Statistical principals for clinical trials. *Statistics in Medicine* 1999; **18**:1905–1942.
3. Zelen M. Biostatistical science as a discipline: a look into the future (With discussion). *Biometrics* 1983; **39**:827–837.
4. Chiang CL. What is biostatistics? (With discussion by M. Zelen). *Biometrics* 1985; **41**:771–775.

DOI: 10.1002/sim