THE APPLIED STATISTICAL SCIENTIST IN A HIGH-PROFILE ACADEMIC ENVIRONMENT

Geert Molenberghs
International Institute for Biostatistics and statistical Bioinformatics
Universiteit Hasselt & Katholieke Universiteit Leuven, Belgium
geert.molenberghs@uhasselt.be

We describe the importance of education in the high-profile, interdisciplinary academic environment facing the applied statistical scientist. Specific attention will be given to the context of biometry and medical statistics. The role can be played by international professional organizations is discussed.

INTRODUCTION

Our discipline is known under the names statistics, biometry, biostatistics, etc. We view all of these names as directed at the same general enterprise of the use and development of statistical theory and methods to address design, analysis, and interpretation of information in the biological sciences. Whatever it is called, our profession clearly is at a critical juncture. Our work has always been and continues to be of great importance in the conduct of scientific investigations in agriculture, life sciences in general, ecology, forestry, medicine, public health, and host of other endeavors. However, with perhaps a few exceptions, our field overall has not enjoyed the recognition it richly deserves as a fundamental cog in the wheel of scientific inquiry, neither among subject-matter scientists nor government officials nor the general public. This has never been more pressing than in current times, with advances in technology that have allowed amazing amounts of information to be collected and stored and with an increasingly complex health-care and public health landscape associated with dizzying arrays of new pharmaceutical products and medical procedures and growing interest among the public in assessing their risks and benefits.

Personal experiences of many of us, collaborating with scientists in agriculture and life sciences and with physicians, and observing the role of our discipline in public life, have often been very rewarding but have sometimes been disheartening. Although some investigators have greatly appreciated our contributions, others tend to view statisticians as, at best, "staff" who know how to run software, "secretaries that are able to count." Our profession tends to attract a number of individuals with an attitude that is, with the best of intentions, too service oriented. Some of us have been told that "it is not our policy to have statisticians as co-authors," by horticulturalists, entomologists, and crop scientists (after designing the study, analyzing the data, and writing the statistical section of the journal article); and have argued with physician colleagues about issues such as adjusting for post-randomization covariates, using longitudinal data methods to analyze longitudinal data (rather than comparing means at each time point separately: "we could never get that kind of analysis published"), and so on. In many of these situations, our expertise has been all but ignored as irrelevant. In the media, it is rare to see a statistician interviewed to explain what are essentially statistical issues; the media would rather solicit opinions from "the real scientists," resulting in pronouncements from, for example, physicians on popular news shows that are inaccurate or downright misleading. The impressions of many non-statistician acquaintances about the role of statistics in public policy and scientific inquiry range from woefully naive to complete ignorance. This scenario is not one of success and certainly not one that will lead to continuity.

One very important way for a statistician to counteract this attitude is by becoming an expert, or at least quite knowledgeable, in the subject matter area within which he or she is working. By combining statistical reasoning with knowledge of the real scientific problems, statisticians can and have made high profile contributions both to the science and to the policy. Important contemporary examples are Fred Mosteller, John Tukey, Sir David Cox, Lincoln Moses (incidentally the first four winners of the Marvin Zelen Leadership Award), Nick Day, Richard Peto, Klaus Dietz, and Martin Gardner. Further, we have to teach our colleagues from substantive areas what our contributions are to the big picture. This needs to be done every time we start working with a new colleague. It should never be taken for granted. Applied statistics is, more often than not, an acquired taste to those we work with. Those to whom it is love at first sight are

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

likely to be found in the profession already. Related to this, we need to become better, as a profession, at explaining the statistical methods we use in less technical language.

The *International Biometric Society* being the learned Society dedicated to our field, we believe we are in a unique position to address the challenge of enhancing understanding of and appreciation for our contributions and expertise among scientists and the public alike and have a greater role in public discourse where interpretation of information is a key aspect. Never has this been more important than with the explosion of data collection in areas like genomics, medical imaging, environmental and bio-terrorism surveillance; with the increasing reliance on sophisticated mathematical modeling to explain biological phenomena; and with the increasing public focus on displaying "statistics" at every turn and disseminating results of studies such as those on mammography and hormone replacement therapy. Computer scientists, applied mathematicians, physicists, and others are claiming these areas, in many cases "reinventing the wheel" or, worse, using inappropriate approaches. It is our belief that we have a fundamental responsibility as a Society to undertake this challenge. We believe that all statistically oriented societies, not just IBS, have tended in the past to look inward rather than outward, and we have neglected to capitalize sufficiently on the potential role we could have in educating domain scientists and the public on the importance of our discipline in all aspects of modern life. At the same time, we are convinced that issues affecting the future of the workings of the Society and the future of the profession are inextricably linked.

THE CASE OF THE INTERNATIONAL BIOMETRIC SOCIETY

To get a feel for our case study, consider some data on the International Biometric Society. The membership of the Society, parent society of the journals *Biometrics* and *Journal of the Agricultural, Biological, and Environmental Statistics*, as well as of the biennial *International Biometric Conference*, amounts to roughly 6000 professionals across all continents, organized into 19 Regions and 19 Groups. Further, there are networks, including Sub-Saharan Africa, the Channel Network (France, the Netherlands, Belgium, the United Kingdom, Ireland), the Central European Network (Germany, Austria, Switzerland, Poland), etc. The Region and Group structure testifies to the Society's confederate structure. Our Society, while united in our professional goals, exhibits rich diversity in language and culture. About 3400 members are native English speaking, with 900 having German, 400 Spanish, and 350 French as their native language. Around 1000 speak another language.

A FORGOTTEN PUBLICATION…

In 1948, in the early days of the International Biometric Society (formed in 1947), Chester Bliss edited a proceedings volume entitled Biometrical Clinic on Entomological Problems. This publication came to the attention of Timothy G. Gregoire, past President of ENAR and based at Yale, as it was part of the inheritance of Lewis R. Grosenbaugh, Yale School of Forestry and Environmental Studies graduate, who spent virtually his entire career with the U.S. Forest Service. Grosenbaugh passed away in April 2003 and the text is now on display at Yale.

While over 50 years old, the message is remarkably fresh and relevant, as we can read from the foreword:

> The advances in biometry have developed through the close cooperation of biologists confronted with problems and of statisticians who develop methods for solving them. Both have gained from this collaboration. One medium for maintaining contact is the "biometrical clinic," in which questions are asked by the biologist and answered informally by the statistician. The meeting recorded here followed this pattern. Even when the answers can be found in the textbooks or in scientific journals, the method which is most relevant to a specific problem may not be apparent to the uninitiated. If a question cannot be answered, its asking may guide the statistician into a new and interesting field of study. Hence the session is of interest to both the biologist and the statistician.

This text provides food for reflection, especially at times where we are confronted, in addition to our traditional collaborators in agricultural, environmental, medical, biological, and

epidemiological sciences, with colleagues from various new fields including molecular biology, bioinformatics, genomics, proteomics, even from security biometrics.

Unfortunately, we are so busy trying to get our work published and to keep our CVs up to date, that we are in danger of losing the habit of scientific discussion, among ourselves and with collaborators in substantive fields. Nevertheless, half a century ago, our colleagues realized this type of exchange was vital, leading them to spend time on answering such questions such as, quoting from p.5 of the same publication: "What is the value of replication and randomization if later the data are not analyzed statistically?"

The meeting which is recounted in the proceedings volume was privileged enough to welcome a fine slate of top representatives of our profession, a testimony to the importance given to this type of encounter. The panel included Frank Wilcoxon, Chester I. Bliss, and John W. Tukey. They were convinced that initiatives like this one were beneficial at both the substantive and the biometric end. They knew research, development, and publication had to be done *after* the clinic and not before. The lesson to be learned, perhaps, is that publishing on a real problem, after taking the trouble to deeply familiarize oneself with it before even starting the research, is the correct way. This way of working is more fruitful, in the long run, than coming up with yet another epsilon increment over an existing method, hoping that one day someone will be kind enough to use it. While, luckily, the finest of our contemporaries still practise this art, the "publish or perish" culture tends to get in the way.

The Bliss proceedings volume teaches us how useful it is to understand our own history, especially in the light of the issues we currently face, as sketched in the Introduction.

Having returned to the early days of the International Biometric Society, let us explore this period in somewhat more detail, and draw some lessons for the current time. But first, let us reflect on the very interdisciplinary nature of our field.

AN INTERDISCIPLINARY SCIENCE

Two decades ago, "pure science" was considered "better science." The best students went on to pure mathematics, with the applied branches left to the others. Important breakthroughs in the last decade, primarily in the field of molecular biology, have made it increasingly clear that the future is bright for interdisciplinary and multidisciplinary efforts and that academic institutions should prepare themselves for this trend. A nice example is the *Bio-X* initiative in Stanford, a new entity in a flashy building, encompassing molecular biology, life sciences, biosciences, or whatever we want to call them

It would be natural to think that, by the very nature of our profession, we have a competitive advantage. Not only do we contribute significantly in a scientific way, as evidenced by the large number of journal articles and books in statistical genetics, bioinformatics, and computational biology (Halloran & Geisser, 1999; Balding, Bishop & Cannings, 2001; McLachlan, Do & Ambroise, 2004; Ting Lee 2004), but we also have a long tradition of collaboration across disciplines, with *Biometrika* being century-old (and yes, originally it was an applied journal), and our own *Biometrics* having half a century of tradition as well. However, while Stanford's Bio-X banner rightly puts biology in the center, surrounded by partner sciences, biometry or biostatistics is completely absent. The argument is that we are embraced by the term "information sciences," but other sciences, such as medicine, are not put under an umbrella term.

We have to remain more vigilant than ever, and appropriately fight for our position, not for the sake of merely perpetuating our profession. If we do not serve a purpose, let us disappear! However, we are deeply convinced that we can have a profound beneficial impact on the biosciences. But we will have to live with the quadrature of the circle: we want to be united with our scientific partners in agriculture, biology, and medicine on the one hand, while on the other to maintain links with our colleagues from other statistical branches, whether theoretical or applied. Insulation is lethal, examples abound. A benign effect of insulation is that we use the same name to different things, depending on our area of application. Think, for example, about such terms as "conditional" and "mixed." Another mild effect is that different names are used for the same thing, such as, for example, variance components (Searle, Casella & McCulloch, 1992), mixed models (Brown and Prescott 1999), repeated measures (Vonesh & Chinchilli, 1997), longitudinal data (Verbeke & Molenberghs, 2000, Diggle et al., 2002), hierarchical models, multilevel models

(Goldstein, 1995), growth curves (Kshirsagar & Smith, 1995), random coefficients (Longford, 1993), etc. Naively looking at the titles of these volumes, it looks like they all deal with different topics but, in spite of such differences, there is a strong common denominator. More worrisome is the relative insulation within which developments take place. In this sense, declaring a specific technique within statistics to be a separate profession, is dangerous. This holds, for example, for classification, sensitivity analysis, and statistical modeling. In principle, there is nothing against scientific meetings or societies with a narrowly defined remit, but the communication lines with the professional groups at large have to remain open.

The current trend points in the reverse direction. Rather than splitting the profession ever further into subgroups, lines between existing subgroups start to blur. In some instances, the lines between biometricians and biologists fade out, or between biostatisticians and clinical trialists. Working in such an environment is challenging but, if we are ready to look in various directions at the same time, it can be very rewarding. While individual professionals and academic institutions (e.g., the Bio-X initiative) are readying themselves for the future, also scientific societies will have to formulate appropriate plans. We will return to this point.

SOME IBS HISTORY AND LESSIONS TO BE LEARNED

In its Volume 106, No. 2757, published October 31, 1947, *Science* reported: "International Biometric Society Formed at Woods Hole Conference." The mission statement of the Society is:

> The Biometric Society is an international society for the advancement of quantitative biological science through the development of quantitative theories and the application, development and dissemination of effective mathematical and statistical techniques. To this end, the society welcomes to membership biologists, mathematicians, statisticians and others interested in applying similar techniques.

The first slate of international officers included a number of now famous colleagues. Sir Ronald A. Fisher was the first President, and P.C. Mahalanobis was Vice-President. Chester I. Bliss served as the first Secretary. At its formation, the Society comprised nine Regions: Australian (now Australasian), Benelux, British, ENAR, French, Indian, Russian, Scandinavian, and WNAR. The initial Society reflected the political situation of the world, so soon after the Second World War. Presidents following on Sir Ronald Fisher include William Cochran, E. Cornish, Chester Bliss, David Finney, Gertrude Cox, Peter Armitage, C.R. Rao, John Nelder and, more recently, Jonas Ellenberg, Niels Keiding, Byron Morgan, Norman Breslow, and Tom Louis.

Are times different now? We always like to believe they are. However, we observe that younger colleagues are less inclined to become a member of a society and remain faithful and loyal to it for the rest of their careers. There is somewhat of a 'shopping' or even 'zapping' culture. Indeed, within thirty minutes we can check the publication status of our article on one society's web pages, download a document from another site, and check a third society's site for details on an upcoming conference. This is undeniably convenient and it is a commodity that hopefully will stay. However, it does present challenges to our societies' structure. Because of the said ease of access to conferences and journals, perhaps on a pay-per-view basis rather than as subscriber, researchers feel less of a need to be a society member. However, we believe that both researchers and societies need to think carefully about the current situation and the future.

First, while access to (society) is easy and virtually universal, there is the non-trivial task of creating the journal. This involves the peer review system on the one hand and the production process on the other. I believe that there is strong value in having independent society journals, to protect scientific integrity and independence as well as to ensure quality of the entire review and production cycle. This cannot happen without strong, well organized, and viable societies. The fact that journals are available on-line, sometimes exclusively so, should not imply that scientific, peer-review, linguistic, and copy-editing standard should lower. Rather, all of these should be implemented with the highest levels of rigor.

Second, societies have to recognize that our times, especially for the biomedical researcher, the biostatistician, the medical statistician, and the epidemiologist is above clear boundaries between the various professions. This implies that many scientists will consider more

than one society of interest to them. Rather than having a single societal home, many people adhere to a multitude of scientific societies, but then maybe without the absolute loyalty from the past. In a way societies follow this trend in as far as there nowadays is a patchwork of associations that cater similar, overlapping audiences. For example, the biostatistician and related scientist may affiliate himself/herself with the International Biometric Society, the International Society for Clinical Biostatistics, the Drug Information Association, the Statisticians in the Pharmaceutical Industry, the Society for Clinical Trials, the American Statistical Association, and the Royal Statistical Society, to name but a few. One problem is that it is currently virtually impossible, for budgetary reasons, to become full member of all of these societies. Hence, societies are urged to collaborate in the face of this situation, and propose novel concepts, such as package deals etc.

EDUCATION

Arguably, the most important area of activity for the future is education. This was already hinted upon in the Introduction. We need to invest in facilitating adequate training, both of candidate statisticians as of colleagues in fields with which we collaborate. Educational needs are different in different parts of the world. While ample master and PhD level education in biometry and related fields is available in a number of Anglo-Saxon countries, and some programs exist in some other countries, including in Western Europe, still great needs exist in various parts of the world. Thus, scientific societies ought to be instrumental in providing education at a wide variety of levels, including: short courses and satellite workshops at international conferences, at Regional conferences, and at other scientific events in which we are or can become a partner; Summer or Winter Schools; templates for partial or full graduate programs, facilitating exchange of instructors, etc. One might consider drawing up a template program, based on successful experiences from various locations around the globe. Such a template could consist of a minimal core together with a large number of suggestions for other mandatory and optional courses. In addition to this, successful instructors could be asked to make their course notes available, for use and/or adaptation by local instructors. Such a template could be accompanied by full syllabi for courses, course notes, exercises, standard tests, etc. An important component of this and other educational initiatives is the use of Distance Learning methods (online courses, access to free educational materials, software). Such methods could either replace or supplement on-site teaching.

Many societies are providing professional development services to its members, while others are relatively absent in this niche. Large national societies, such as the Royal Statistical Society and the American Statistical Association take such initiatives very seriously. Inter-society working parties may be set up to provide professional development services, in the spirit of what is said in the previous section.

CONCLUSION

The most important area of activity for the future is education. We need to invest in facilitating adequate training, of candidate statisticians and colleagues in fields with which we collaborate. Educational needs are different in different parts of the world. While ample master and PhD level education in biometry and related fields is available in Anglo-Saxon countries, and some programs exist in some others, still great needs exist in various parts of the world. Thus, scientific societies ought to be instrumental in providing education at a wide variety of levels, including: short courses and satellite workshops at conferences, and at other scientific events in which we are or can become a partner; Summer or Winter Schools; templates for partial or full graduate programs, facilitating exchange of instructors, etc. Such templates could consist of a minimal core together with a large number of suggestions for other mandatory and optional courses. In addition, successful instructors could be asked to make their course notes available, for use and/or adaptation by local instructors. Templates could be accompanied by full syllabi for courses, course notes, exercises, standard tests, etc. The use of Distance Learning methods is essential. Such methods could either replace or supplement on-site teaching. Many societies are providing professional development services to its members, while others are relatively absent in this niche. Inter-society working parties may be set up to provide professional development services, in the spirit of what is said in the previous section.

REFERENCES

Balding, D. J., Bishop, M., & Cannings, C. (2001). *Handbook of Statistical Genetics*. New York: John Wiley & Sons.

Bliss, C. I. (1948). Biometrical clinic of entomological problems. *Proceedings of a joint meeting of the American Association of Economic Entomologists and he Biometric Society, Eastern North American Region.*

Brown, H., & Prescott, R. (1999). *Applied Mixed Models in Medicine.* New York: John Wiley & Sons.

Diggle, P. J., Heagerty, P. J., Liang, K. Y., & Zeger, S. L. (2002). *Analysis of Longitudinal Data*. (2nd edition). Oxford: Oxford University Press.

Goldstein, H. (1995). *Multilevel Statistical Models.* London: Edward Arnold.

Halloran, M. E., & Geisser, S. (1999). *Statistics in Genetics.* New York: Springer.

Kshirsagar, A. M. & Smith, W. B. (1995). *Growth Curves.* New York: Marcel Dekker.

Longford, N. T. (1993). *Random Coefficient Models.* Oxford: Oxford University Press.

McLachlan, G. J., Do, K.-A., & Ambroise, C. (2004). *Analyzing Microarray Gene Expression Data*. New York: John Wiley & Sons.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance Components.* New York: John Wiley & Sons.

Ting Lee, M.-L. (2004). *Analysis of Microarray Gene Expression Data*. Boston: Kluwer.

Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York: Springer.

Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and Non-linear Models for the Analysis of Repeated Measurements.* Basel: Marcel Dekker.