Training the Next Generation of Biostatisticians

David L. DeMets, Dan Anbar, William Fairweather, Thomas A. Louis, and Robert T. O'Neill

Biostatisticians have become an integral part of medical research for the design and analysis of laboratory, epidemiology, and clinical studies. The demand for this expertise has increased over the past several decades, while the supply has remained relatively constant. The skills required to be a successful statistical collaborator go beyond the traditional statistics courses taught in graduate school. We describe the challenges facing those training the next generation of biostatisticians.

KEY WORDS: Biostatistics; Training.

1. INTRODUCTION

Biostatistics as a recognized discipline is less than 50 years old, having started in the late 1940s with the growth of systematic medical research in the United Kingdom and the United States. The development of biostatistics has been particularly noticeable over the past 30 years with the increasing use of controlled clinical trials for the evaluation of new drugs, which has been stimulated by federal regulations requiring evidence for safety and efficacy. This growth has been seen in academic teaching and research environments, in the pharmaceutical industry, and in federal regulatory and research agencies such as the Food and Drug Administration (FDA) and the National Institutes of Health (NIH). The NIH in particular has been a leading institution in this development. Recently the explosion of research and development in biotechnology has opened new areas for the application and development of biostatistics. There is also increasing emphasis on controlled trials of medical devices to establish their safety and efficacy. The field of statistics, which focuses on medical, biological, and environmental sciences, has also developed an international professional society, the International Biometric Society, with a membership of over 5,000. Biostatisticians focus mostly, but not exclusively, on the medical sciences.

The biostatistician has become an integral member of the medical research team over the past decades, and it is likely that this trend will continue. Until recently, with rare exception the biostatistician's role was to design the statistical aspects of the experiment or study, gather and organize the data, and analyze the results. This role has not changed. However, with the availability of powerful computer hardware and user-friendly computer software, which enables nonstatisticians to easily perform sophisticated statistical analyses, statisticians must adjust their focus and find ways to take advantage of these tools while avoiding some of the inherent dangers. The statistician who waits for the customer to come to the office or the classroom and who would dispense only textbook advice will fail to meet the new challenges in medical research.

Funds for research, within the public or private sector, are limited. Good experimental design and analyses make proposals better and more competitive in both industry and academia. Grant proposals submitted with design flaws or poor statistical support to the NIH or other funding agencies are unlikely to be funded. Analyses of completed projects that stand up to peer review do not go unnoticed in this competitive research enterprise. Government biostatisticians affect the spending of billions of tax dollars, and in the case of the FDA, biostatisticians assess the outcome of research to determine whether the products should be released to the public or removed from public consumption. For regulated products, the law requires that these actions be based on scientific studies and relevant data. In industry, 15% of the sales of a pharmaceutical company are invested in research and development on which the biostatistician can have a major impact. Effective design and analyses mean better studies, faster product licensing if the product is effective, and thus larger corporate profits as well as public benefit. Furthermore, through their training, statisticians acquire a systematic approach to planning and problem solving. This skill is not limited to study design and analysis but is basic to scientific leadership.

More and more medical and biological research where statistical issues are encountered is done outside of academia. Consequently, the gap is widening between what the traditional approach to training prepares the statistician for and what is encountered upon starting to practice statistics. This poses a challenge to academicians and those in government, and industry at a time when financial resources as well as good candidates are scarce.

As we look toward training the next generation of biostatisticians, we must face several challenges in order to ensure the health of our profession, which is a critical discipline in biomedical research. These challenges include the growing demand for biostatistics, the decreasing supply of biostatisticians, the shrinking training support, the need to reshape curriculums, and the need to develop a closer working relationship among academia, government and industry. Responses to these challenges must be found by rethinking our approaches to training. These and other issues are discussed in this article.

2. ESTIMATING SUPPLY AND DEMAND

One of the immediate problems of training the next generation is that, as a biostatistics profession, we do

David L. DeMets is Professor and Chair, Department of Biostatistics, University of Wisconsin-Madison Medical School, Madison, WI 53792. Dan Anbar is with the Biostatistics Department, Schering-Plough Corporation, Kenilworth, NJ 07033. William Fairweather is Branch Chief, Biometrics, U.S. Food and Drug Administration, Rockville, MD 20857. Thomas A. Louis is Professor and Head, Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455. Robert T. O'Neill is Acting Director, Office of Epidemiology and Biostatistics, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Rockville, MD 20857.

not have an official or formal data base to monitor the supply and demand for M.S. and Ph.D. level biostatisticians. However, the National Research Council (1987) collects data on the number of doctorates in the physical and biological sciences, including mathematical statistics and biometrics/biostatistics. Since 1980, U.S. universities awarded a yearly average of approximately 150 doctorates in mathematical statistics and 40–50 doctorates in biometrics/biostatistics. Approximately 50% of the candidates in mathematical statistics and 80% of those in biostatistics currently are U.S. residents. Each year, *Amstat News*, the official newsletter of the American Statistical Association, publishes a listing of all statistics and biostatistics departments in North America. According to *Amstat News*, approximately 40 biostatistics or biometry departments exist with Ph.D. programs. In addition, there are perhaps another 10–12 statistics programs with some emphasis in training statisticians with interests in biostatistics. Each year these listings provide the number of M.S. and Ph.D. degrees awarded. Based on this information, it appears that, since 1985, this collection of departments produced annually about 50 Ph.D. level biostatisticians, in agreement with the National Research Council data.

From 1985 to 1993, *Amstat News* position advertisements for biostatisticians increased from 100 to almost 200. However, it is difficult to be sure that some of the advertisements were not for the same position. Industrial positions are likely underrepresented in these advertisements since recruitment is often done through professional recruiting agencies. Imperfect as this is, it seems safe to argue that the number of positions needed exceeds the number produced and that the gap is widening. Roughly 45% of the positions are in academia, 45% in industry, and 10% in government. It is likely that the number of positions in industry will continue to increase with coming changes in the regulatory environment in Europe and the anticipated increase of statistical input at the FDA as a result of user fees. Demand will increase in academia and government as well because of greater scrutiny of health care costs and the use of procedures and devices.

Most academic training programs in biostatistics recruit students from undergraduate or graduate programs in mathematics or computer science with an occasional student from the social or biological sciences. Based on experiences at the University of Wisconsin and the University of Minnesota, we estimate that about 80% of the biostatistical students come from the former background. However, the pool of American undergraduate students majoring in mathematics is decreasing. Furthermore, fewer of those are going on to graduate school in statistics or biostatistics. A quick survey of a few institutions indicates that 50% to 70% of statistics/biostatistics students in graduate programs at the Ph.D. level are U.S. trained as undergraduates. Thus, at a time when we need to produce more biostatisticians, the talent pool in the United States from which to recruit appears to be decreasing.

Graduate programs in the biological and physical sciences in general have an increasing number of foreign students filling the void created by the decreasing number of American students. Biostatistics is experiencing a similar phenomenon. This greater dependence on foreign students to meet the demand may require reevaluation of eligibility for federally funded training programs and for the nature of our training in communication skills. These will be discussed later on.

As already indicated, these preliminary assessments are based on very soft data. We are unaware of other hard data specific to biostatistics. As a profession that prides itself in the design of studies, data collection, and analysis of data, it is ironic that we do not have very good information on the current and projected demand, on the supply of students trained each year, the characteristics of their training, and some follow-up on the employment of those entering the job market each year.

3. TRAINING SUPPORT

There are three general sources of funding and financial support for biostatistics graduate students: university, government, and industry. The university sponsoring the graduate program may provide departments with support for teaching assistantships (TA's) or fellowships, but this support is dwindling. Generally, fellowships are less frequent or limited in the number of years of support. TA support may cover duties ranging from grading papers to conducting discussions or help sessions to actual classroom lectures. Both private and public institutions are undergoing fiscal constraints or cutbacks. Thus, biostatistics departments will be subject to these general fiscal pressures and are not likely to increase the number of TA's; in fact, they may be facing a decrease.

Over the past 30 years, many students have been supported by NIH training grants. Until 20 years ago, NIH training grants for biostatistics, which covered not only student stipends but also salaries for faculty, were awarded competitively to biostatistics departments. In the early 1970s, the NIH cut back its training grants, and many biostatistics departments scrambled to find alternate sources of support for faculty and students. Many relied on contracts or grants for statistical coordinating centers that provide support to clinical trials. Today many large training programs rely heavily on this source of support. During the 1980s, individual disease-oriented NIH institutes began to support training of biostatisticians where departments had strong ties to medical research in that disease, such as cancer or cardiology. With the small increases over the past several years in the NIH budget, this source of support is also not likely to increase without some special attention. Furthermore, not all institutes at NIH sponsor training programs in biostatistics, so an individual department may be limited by the specific disease-oriented research conducted in that institution.

Recent reviews of some of the authors and their personal experience with the NIH review process for biostatistics training grant applications cause concern. There is no standing review committee specifically for biostatistics applications. Rather, these applications are handled by committees that review a wide range of disciplines. There is no guarantee that a training grant for biostatisticians will ever be seen or primarily reviewed by a biostatistician. Another aspect of the review process is the criteria for successful performance of a training program. The criteria include the number of publications, the number of NIH grants, the amount of funding, and the institution of employment of former students. By these criteria, employment in an academic research institution implicitly counts more than nonacademic research alternatives. The inference is that a highly qualified student going to work for a federal or state agency (e.g., the NIH, the FDA, State Departments of Health) or for private industry (consulting firms, pharmaceutical companies) would not be given as much credit as the student entering an academic research institution. To favor those entering academic research institutions is misguided and wrong.

Other sources of support come from limited biostatistics research programs sponsored again by one of the disease-oriented institutes. The competition for NIH research grants is intense and budgets are frequently cut so that support of research assistants (RA's) is limited or threatened. Furthermore, these RA positions are intended to conduct the proposed research in the awarded grant and are not specifically for the training of a student, even though the two goals may be quite compatible in many situations.

Industry has traditionally provided indirect support for the training of students through research contracts with biostatistics departments that provide support for clinical trials or basic science research, similar to NIH-funded trials. More recently, some companies have developed summer internship programs that offer students the opportunity to gain experience and to become familiar with careers in industry.

4. WORK ENVIRONMENT AND SKILLS

In academia, government, and industry, biostatisticians now operate in an environment where, in addition to their traditional technical roles, they need to provide scientific and organizational leadership. The biostatistician may have responsibility for teaching, consulting, scientific collaboration, and statistical research. For a consulting or collaborating biostatistician, the largest single component of time is spent interacting with scientists, collaborating on the conceptualization of a problem or the design, conduct, and analysis of a study, refining the scientific question, and discussing the results of analyses or the drafting of a scientific paper. The key elements are interpersonal skills including listening, proposing, and negotiating. A biostatistician must have the ability to see both the big picture and the technical details, and the ability to present issues and findings effectively before a critical, possibly nontechnical and multidisciplinary audience.

Once the collaboration is underway, the biostatistician often may be responsible for managing the data or for assuring that the data management is conducted in a manner amenable to data analysis. This may mean supervising a programming staff, data managers, and other clerical staff as well as being responsible for both the hardware and software necessary for either data management or analyses or both. Management skills are important. In addition, knowledge of data collection procedures such as forms design and quality control are also essential. Some students get exposed through work experience or special courses while others never get this opportunity.

The biostatistician often faces the challenge of analyzing messy data with missing items, outliers, and skewness. Increasingly biostatisticians use nonlinear models, and analysis must be summarized through graphics, potentially by tables, and in writing. Communications skills in presenting the analyses and in writing a report are necessary to be effective. To conduct the analyses, the biostatistician may have to learn new statistical methodology and how to apply it to the particular situation. In a few instances, a particular experiment and its resulting data may not lend themselves to standard analyses, and new methodology must be developed.

Thus, biostatisticians spend much of their time communicating and interacting with colleagues and collaborating researchers, defining the research question and interpreting results of analyses. More time is spent on aspects of data management and analyses than on developing new statistical methodology. Knowledge of computer software for analyses and data management are also essential. Given the nature of statistical consulting and collaboration, the biostatistician must be familiar with a wide range of statistical theory and methods. The biostatistician must also have leadership qualities to supervise support staff and to direct and influence the narrow and broad statistical aspects of a research project. It is not sufficient to provide the correct answer to a design or analysis problem if and when asked. The biostatistician must be willing to initiate a conversation, to schedule and run a meeting that addresses a possible problem, present new insight, or propose a possible revised solution. The biostatistician must be able to negotiate an acceptable approach rather than just provide the correct technical response to an inappropriately posed question.

5. CURRENT TRAINING

The biostatistician has a marvelous opportunity to make a meaningful impact on medical and biological research by integrating and employing a variety of technical and nontechnical skills. Knowledge of statistical theory and methods is necessary but not sufficient. Many biostatisticians in their graduate training spend most, if not all, of their time on statistical theory and methods and very little on the critical nontechnical skills such as interaction, communication, and leadership. If we are to prepare the next generation of students for exciting challenges and opportunities, our curriculum needs to change to provide these students with a broader background.

The typical Ph.D. program focuses three to four years of course work on a range of topics ranging from mathematical statistics and probability to experimental design, data analysis, and computing. The master's curriculum usually includes a two-year subset of these formal courses. Some programs may require a statistical consulting course. Qualifying exams may be based on theory and methods, the mix depending on the training program. Students may gain experience in consulting or collaboration in some training programs with a connection to a medical research center. Despite what seems like excellent training, many employers find that new Ph.D.'s need one or two years of on-the-job training before becoming effective. Many biological disciplines expect that new Ph.D.'s will spend one or more years in postdoctorate training, making the transition from student to independent scientist. This has not been a tradition in biostatistics although some departments prefer to hire faculty with postdoctorate experience and may even allow biostatisticians to spend a year or two as a postdoctorate before converting them to faculty or staff positions. The lack of a postdoctorate tradition is due in part to the shortage of biostatisticians and of funds for such training.

Another challenge in training is that over the past 20 years (especially the last 10 years) the amount of statistical research and new methodology has grown substantially. This explosion creates more technical material to be put into an already full curriculum. If new emphasis is placed on training students in communication skills and leadership in order to be more effective as consultants and collaborators, either something has to be eliminated or reduced, or the augmented requirements and extended time for completing the M.A. or Ph.D. will become intolerable for faculty, (potential) students, and sources of financial support.

6. WHERE DO WE GO FROM HERE?

Although there are signs of serious gaps in the training of the current and next generation of biostatisticians, the future potential for this discipline is very bright if we can meet the training challenge. Given the pivotal contributions that biostatistics can bring to biomedical research and policy, we must now begin to meet the challenge so that, by the end of the century, the next generation will be properly trained. We offer some proposals to meet the challenge.

First, obtain better data about supply and demand. We need to work with the International Biometric Society and the Biometrics Section of the American Statistical Association to keep a current data base of the number of students applying for graduate training, relevant demographics, the number of students in biostatistics/biometrics training programs each year, the number of students graduating, and their initial place of employment. In addition, we need to keep better accounting of employment opportunities, the placement rate, and the length of time required to complete the search.

Second, conduct a market survey of employers that register their vacant positions with *Amstat News* or the International Biometrics Society Office. This survey should reveal out the nature of the position and the type of skills employers believe necessary for a new Ph.D. to be successful.

Third, a survey of recent graduates and those employed for 10 years should be conducted to determine their perceived needs. We can then evaluate whether our training has been complete and thorough or if there are gaps.

Fourth, an integrated effort must be made at the federal level to provide coordinated training support for students. Currently, the sources of federal support from the NIH are fragmented, insufficient, and often unnecessarily restrictive.

Fifth, departments that train biostatisticians must review and modify their curricula to meet changing and increasing demands. Many employers seem satisfied with the technical training in mathematical statistics and methods of analysis but point out deficiencies in communication and leadership. Further, biostatisticians should be coached in presentation skills before small and large groups and be taught to utilize slides and transparencies effectively. These skills will not be learned spontaneously or in a short period of time in a typical course. Rather, they must be introduced early in graduate training and practiced over the training period.

Sixth, students should be introduced to real studies and investigators early in their graduate training to develop a sense of application of methods and their limitations. In order to better communicate with fellow scientists, biostatisticians must gain some science background and have a research curiosity. We need to make better use of our research projects as training opportunities for data collection procedures, quality control, meetings of collaborators, conference calls, and the like. Of course, this means better planning, more integration, and increased effort on the part of faculty as well as students.

Seventh, government, industry, academia, and statistical societies must work in closer partnership. Teaching faculty need to take sabbaticals or consult and collaborate more with colleagues in industry and government. In addition, government and industry statisticians have a lot of experience to share in the classroom. For academia, this interaction provides an additional stimulation to research problems as well as an opportunity to apply new methods and have a more immediate impact of their methodologic research. In addition, students exposed to interaction with industry or government will have better insight and opportunity for employment. Industry and government will also benefit from this collaboration by having the opportunity to impact on research directions in academia, to influence curriculum to meet their needs, and to receive better trained employees and faculty consultants. Government and industry statisticians would have more rapid access to new research and be able to collaborate on research issues they face in their environment.

Finally, our profession must do a better job of promoting ourselves to society, policy makers, and senior-level management in industry and government. Potential colleagues often do not understand or appreciate the scope of our possible contribution and the organizational and financial prerequisite for success. Too often we are viewed and treated as technicians and not as scientific colleagues. That failure lies with us. We must get actively involved and provide leadership.

7. CONCLUDING REMARKS

Biostatistics is an exciting and rewarding profession. It will continue to be if we care for and nurture it. In particular, we must pay attention to the training of the next generation and prepare society for us. The goal is to develop a biostatistical scientist engaged in biomedical research and policy. The next generation of academic, government, and industry statisticians will be able to participate in the leadership of public health improvement depending on their ability to make their ideas understood and to convince others of their worth. To have professionally satisfying options, universities must respond to the growing need for leaders and partners in applied scientific research in addition to the traditional need to develop technical experts in statistical theory and methods. If we meet this challenge, we will succeed in having an impact in a critical time of health care and biological research. If we as a profession do not succeed, we will be left out and those individuals with computer literacy will take our place. All of the issues raised can be addressed if we have enthusiasm, commitment, and persistence. We must return to the next generation the investment the previous generation made in us.

[*Received February 1994. Revised June 1994.*]

REFERENCE

National Research Council (1987), *Summary Report 1987—Doctorate Recipients From United States Universities*, Washington, DC: Academic Press.