WILEY InterScience®
DISCOVER SOMETHING GREAT

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

Training of the next generation of biostatisticians: A call to action in the U.S.

David L. DeMets[1,*,†], Gary Stormo[2], Michael Boehnke[3], Thomas A. Louis[4], Jeremy Taylor[3] and Dennis Dixon[5]

[1]*Department of Biostatistics and Medical Informatics, University of Wisconsin, School of Medicine & Public Health, 600 Highland Avenue, Box 4675, Madison, WI 53792, U.S.A.*
[2]*Department of Genetics, Washington University School of Medicine, St. Louis, MO, U.S.A.*
[3]*Department of Biostatistics, University of Michigan, School of Public Health, Ann Arbor, MI 48109, U.S.A.*
[4]*Biostatistics Department, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, U.S.A.*
[5]*National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, U.S.A.*

SUMMARY

Two workshops (2001, 2003) were held by the National Institutes of Health (NIH) to examine the need to train more biostatisticians in the U.S. to meet the increasing opportunities in the biomedical research enterprise. The supply of new PhD graduates in biostatistics in the U.S. has been relatively steady for the past two decades while the demand has increased dramatically. These workshops concluded that a renewed effort must be made in the U.S., led in part by the NIH, to add to and expand the existing training programs to increase the supply. This article summarizes those two workshops and their recommendations. Some progress has been made through a new biostatistics training program with emphasis in bioinformatics sponsored by the National Institute of General Medical Sciences (NIGMS). Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: biostatistics; bioinformatics; training program; collaboration

INTRODUCTION

Biostatistics is a fundamental scientific component of biomedical, public health and health services research. Traditional and emerging methodological and application areas include clinical trials,

*Correspondence to: David L. DeMets, Department of Biostatistics and Medical Informatics, University of Wisconsin, School of Medicine & Public Health, 600 Highland Avenue, Box 4675, Madison, WI 53792, U.S.A.
†E-mail: demets@biostat.wisc.edu

observational studies, imaging and genomics. This next decade may be the most exciting era in the more than 50 year history of biostatistics in medical and biomedical research. The opportunity to contribute to the identification of new genes which contribute to disease, the development of gene transfer strategies and gene therapy clinical trials, and the development of novel pharmaceuticals or biologics for the treatment or prevention of disease is enormous. However, the shortage of biostatisticians with training in genetics, molecular biology, and bioinformatics is often rate limiting for novel research and discovery for academia as well as for biotechnology and pharmaceutical companies.

The current demand for biostatisticians in the U.S. far exceeds the supply and the gap is expected to continue to widen due in part to the key role biostatistics now plays in the analysis of genomics, proteomics and related research arenas. This under supply results from a combination of the burgeoning need and a relatively flat rate of graduating doctoral and masters level biostatisticians in the U.S. In addition, the number of domestic undergraduates entering graduate programs in biostatistics has not increased and in fact appears to have decreased. Current training programs in biostatistics have not attracted these quantitatively talented undergraduates to take advantage of the many exciting opportunities. A large proportion of the currently senior biostatisticians in the U.S. were trained in the late 1960s and early 1970s and were supported during the early stages of their careers by National Institutes of Health (NIH) training initiatives. With this cohort approaching retirement over the next decade and the tremendous needs and opportunities, both workshops recommended that now is the time for the NIH and the private sector to re-invest in the training of the next generation of biostatisticians within the U.S. The situation world wide is probably similar to that experienced within the U.S. although these workshops did not address this.

This next generation will need to be trained not only in traditional statistical theory and methods, including special topics such as clinical trials and survival analysis, but also in bioinformatics and basic biology as well as 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.

GROWTH IN MARKET DEMAND

In the last 20 years, the demand for biostatisticians has grown dramatically; the role of biostatisticians and biostatistics in the design, conduct and analysis of biomedical research has continued to expand. Areas of increased activity include epidemiological studies, clinical trials, health services assessments, basic laboratory research, biomedical imaging and most recently, genetics and genomics. In addition, neuroscience and computational biology are growth areas.

The amount of data being massed alone in all areas of biological science requires new thinking and training as to how to think about very large data problems. Now more than even there is a market demand for scientifically (i.e. subject matter) trained data analysts (i.e. biostatisticians) to lead the way in making correct inference and scientific study of these large data problems.

In 1998, DeMets and co-authors reviewed the number of positions advertised in the *Amstat News* (the official newsletter of the American Statistical Association) during the years 1990–1994 [1]. Most academic and government research PhD level positions are advertised in this newsletter as are some industry positions. A smaller proportion of MS positions and an even smaller proportion of BS level positions are advertised in this newsletter or its web-based equivalent. Senior or
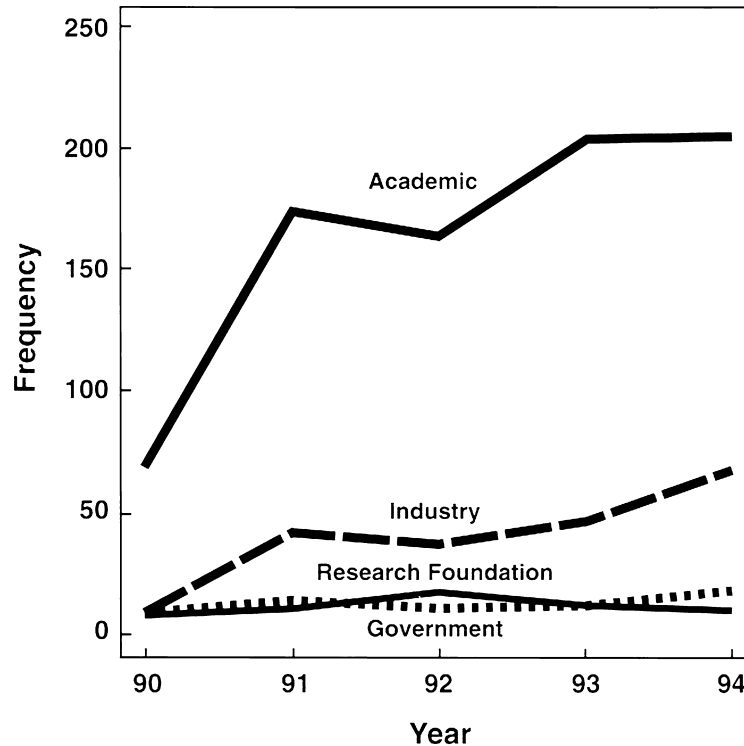
Figure 1. Frequency plot of source of positions *versus* year.

management positions in industry are often placed with recruiting firms and are not advertised openly. Thus, the number of positions listed in *Amstat News* is quite likely to be an underestimate of open positions.

As shown in Figures 1 and 2 [1], the number of advertised positions has continued to increase during the early 1990s. From 1990 to 1994, the number of PhD level positions advertised rose from 60 per year to over 200 per year, most of which were in academia. The number of advertised PhD positions during that period is more than the number of graduates by perhaps a 2 to 1 ratio. In fact, the actual ratio of vacant positions to available candidates is likely to be much larger, and the current ratio even larger due to increasing demand. Unfortunately, we are not aware of any update of these figures in the subsequent decade. However, our experience suggests that the gap between supply and demand may have gotten worse since these figures were first published.

Most biostatistics/statistics departments or biomedical research facilities are having increasing difficulty in filling vacant or new faculty or staff positions. Positions may remain vacant for many months or even years and current biostatistics departments are competing for the same dozen top candidates, not just in the established methodology areas (e.g. survival, longitudinal) but especially in the emerging fields such as microarray data and statistical genetics. As a result, market forces have driven starting salaries up dramatically. Table I [2, 3] displays median salaries taken from surveys for the years 1995–2005. As can be seen, starting salaries have increased substantially during this 10-year period. Significant increments are likely to continue as competition for existing and new talent increases.
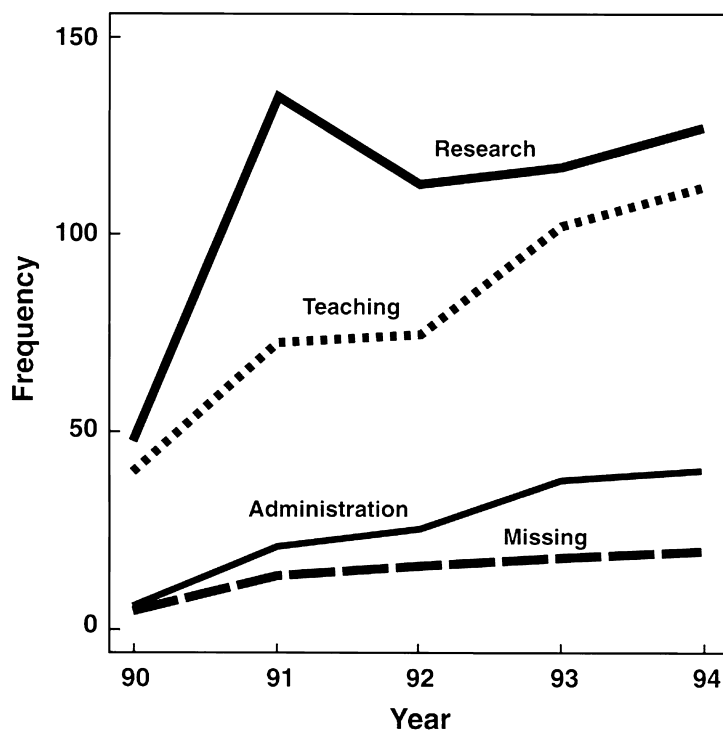
Figure 2. Frequency plot of type of positions *versus* year.

EARLY TRAINING INITIATIVES

With the demand for biostatisticians increasing in the mid 1960s, the NIH launched a large, centrally administered training initiative, funding training grants at many biostatistics and statistics departments. These training grants provided pre-doctoral tuition, pre- and postdoctoral stipends and faculty support for teaching and training. Over the next two decades, this 1960s/1970s cohort became a major influence in biomedical research, with graduates of these training programs assuming leadership positions in the federal government (e.g. at the NIH and the FDA), faculty positions at academic biomedical research institutions and leadership roles on private industry research teams. Many in this cohort chaired departments or managed research groups.

RECENT TRAINING INITIATIVES

In the early 1970s, the NIH terminated the general biostatistics training grants program, believing that the initiative had stimulated the training process and objectives had been met. For the next three decades, NIH-sponsored training grants for biostatistics surfaced in individual institutes. The number of funded pre- and postdoctoral positions is not larger than in the early 1970s and training is largely focused on institute-specific diseases rather than on a broad, biomedical research

Table I. Median salary survey results: biostatistics and other biomedical statistics departments and units (12 months) 1995–2005.

Rank\year in rank	Fall 1995	Fall 1996	Fall 1997	Fall 1998	Fall 1999	Fall 2000	Fall 2001	Fall 2002	Fall 2003	Fall 2004	Fall 2005
Assistant 1–3	$56 072	$57 463	$55 000	$60 000	$61 876	$65 000	$70 000	$78 500	$78 309	$78 640	$83 000
4 or more	$56 000	$57 000	$62 804	$63 860	$67 275	$67 500	$71 450	$77 000	$79 692	$83 408	$87 000
Associate 1–2	$73 091	$72 589	$69 000	$75 232	$80 625	$84 464	$85 608	$78 280	$94 000	$96 056	$100 215
3 or more	64 500	$68 234	$73 840	$76 200	78 749	$86 875	$89 714	$94 212	$98 219	$97 000	$103 334
Full 1–6	$93 067	$93 067	$96 188	$98 094	$100 278	$106 947	$110 401	$114 796	$125 000	$130 500	$147 988
7 or more	$97 516	$98 296	$106 438	117 695	9116 827	$115 885	$125 692	$140 221	$145 712	$148 000	$151 410
Starting Assistant Professors	$50 250	$53 250	$57 000	$59 000	$60 905	$73 217	$67 504	$68 750	$75 000	$79 342	$87 000

relevant education in biostatistics. These disease-focused, biostatistics training grants are reviewed and funded as part of each institute's training programs which are dominated by biology-oriented proposals. Few proposal reviewers are biostatisticians actively involved in the direction of training programs in biostatistics.

Consequently, the goals and impact of biostatistics training grants may not be fully appreciated. In addition, reviewers of disease focused training program proposals generally expect to see a critical mass of a department's faculty involved in research on that disease. Small and mid-sized departments with an excellent training program are unlikely to have such a critical mass, making it hard for them to compete successfully for such grants. This is unfortunate because biostatistics departments with a critical mass of collaborating cross disciplinary biological research can and do contribute meaningfully to the training of biostatisticians.

BIOSTATISTICS GRADUATE PROGRAMS AND STUDENTS

Since the 1970s, few new biostatistics training programs have been developed. A National Science Foundation (NSF) survey [4] found that the number of programs in statistics or biostatistics in the period 1985–1994 increased from 137 to 156, a 14 per cent increase. And, new quantitative science programs such as computer science and informatics are now competing for mathematically skilled graduate students. Schaeffer states that through *Amstat News* online, universities report for 1999–2000, 1100 bachelor's degrees, 1600 master's degrees and 460 doctorates in statistics in general [5].

A survey by Sanderson *et al.* [6] reports that out of the 495 PhD graduates with statistics-related degrees, 76 were in biometrics and biostatistics. The report, *Doctorate Recipients from United States Universities—Summary Report 1999* [7], shows that from 1989 to 1999, the number of PhD graduates in bioengineering and biomedical fields has increased 200 per cent, while the number of PhD graduates in biometrics and biostatistics increased from 46 to 76, a 65 per cent increase. The report lists the number of doctorate recipients by citizenship. Of those who reported their citizenship status, the proportion of non-U.S. citizens with temporary visas for the fields of bioengineering and biomedical, and biometrics and biostatistics were 37 and 56 per cent, respectively. In 2001, a University of Chicago survey indicated that the number of doctoral degrees in biometry/biostatistics was 90, 31 per cent being non-U.S. citizens [7]. Our own experience suggests that these trends continue and that if anything, a decreasing proportion of degrees in these fields are being earned by U.S. citizens.

Directors of education programs are well aware that the number of domestic students applying to graduate school to pursue MS or PhD degrees has not increased and has apparently decreased. Department chairs estimate that a half or more of current students are not U.S. citizens. Competition from other disciplines for domestic students with mathematical talent has increased. For example, computer science is now of interest to a large number of students where the national shortage is also very large. Talented students may pursue mathematically oriented curriculum in business schools, actuarial science, and industrial engineering which are areas that have also increased in opportunity. The interest by high school and college students in mathematical sciences has waned over the past two decades. Foreign students are very talented, they are valued members of our graduate programs and their education is important both for the U.S. and their home countries, but the U.S. cannot afford to rely solely on foreign students to meet its needs in biostatistics. Further, opportunities for foreign students to return to their home countries—particularly China

and India—are increasing, suggesting a likely further erosion of the statistics talent-base in this country.

BIOSTATISTICS POSTDOCTORAL TRAINING

While postdoctoral training has been a traditional step in the education process for most biological sciences, it has not played a similar role in the career of most biostatisticians. Postdoctoral training is a valuable experience for many PhD graduates who desire an academic career, especially those graduating from programs other than Biostatistics. However, market forces make attracting highly qualified students into low paying postdoctoral positions very challenging and the supply of PhD graduates is simply not adequate to fill the available postdoctoral positions.

Given the current demand to broaden a biostatisticians training, 1 or 2 year postdoctoral training should be given new consideration, either supported by NIH funded training grants or industry grants. In addition to consideration for recent PhDs in biostatistics, postdoctoral training should also consider recent PhD graduates in biological sciences or MDs with relevant research interests. Such cross disciplinary postdoctoral training would have the added benefit of encouraging and facilitating cross disciplinary research teams which is paramount for future research success. However, salary levels for post doctoral trainees are far below the salaries of entry level faculty positions, many being tenure track positions, competing for the same talent. Unless this gap is changed, it will continue to be difficult to recruit recent PhD graduates into post doctoral training positions.

CURRICULUM ISSUES

DeMets and co-authors [8] noted several years ago that some curriculum revision is needed to reflect the dramatic expansion of statistical techniques and applications. Students struggle to include all the interesting and relevant courses into their program and still be able to complete their doctorate degree in a reasonable period of time (generally 5–6 years). Most graduate students do not get adequate training in non-technical areas such as consulting, oral and written communication, leadership and management. In order for future generations of biostatisticians to be really effective in collaborating with biological and clinical scientists such skills should be available as part of any biostatistics training program. Increased emphasis on scientific or medical training is also critical, especially in the current genomics era. The next generation of biostatisticians should have a pre- and postdoctoral educational experience quite different from the impressively successful, high-impact 1970s cohort.

The value of exposing students to non-statistical biomedical research is well-recognized. Many academic departments have a formal breadth requirement in a non-statistical field. Also, many students obtain valuable experience working as a research assistant, exposing them to the real world research process including collaboration with non-statistician, biomedical researchers. While such experience is valuable and necessary, it needs to be balanced with the principal goal of training in the theory and application of biostatistical methods. For sure, future students need to know more biology than previous generations and yet acquire knowledge of an ever increasing base of statistical methodology. Achieving this balance is and will continue to be a challenge for academic biostatistics and statistics departments as they assess their current curriculum to meet future needs of students and do so within a reasonable period of time.

In general, we have to think critically about how to reformulate the curriculum for statistical training to balance the need to train students in new areas of statistics while retaining the important aspects of current statistical theory and methodology and yet make sure that students receive formal training as well in specific biological disciplines such as physiology, cellular biology, molecular biology, genetics and neuroscience.

BUDGETARY ISSUES

As NIH budgets became squeezed during the 1980s and 1990s, the level of support for each training grant also became squeezed. The NIH also decided to limit the total amount of tuition that could be charged to a training grant. This limitation can be problematic for institutions with large tuitions, because the shortfall has to be made up from non-federal funds. While this problem is not unique to biostatistics programs, it does place biostatistics at a disadvantage when competing for students with disciplines that have greater available resources. In addition, supplies and expenses for trainees can be inadequate. For example, training grant support for computing resources for a biostatistics graduate student is far less than actual cost.

FUTURE CONSIDERATIONS

While biostatisticians are rightfully proud of past successes and impact over the past 50 years, especially during the past three decades, it is important to ensure that these contributions continue. The 1970s cohort is nearing their retirement years, the current supply remains relatively constant and the demand continues to escalate. While academic departments must assume responsibility and take some leadership, they will require the financial support of the NIH. Unless new training initiatives are launched in the very near future, there will not be adequate time to begin to meet the gap before the 1970s cohort retires. Not meeting this challenge will not only have the effect of slowing down biomedical research progress, especially in the emerging fields such as genomics and bioinformatics, but also will be very detrimental to the profession and the nation.

NIH AD HOC MEETING, SEPTEMBER 2001

Motivated by these needs, an *ad hoc* meeting of biostatistics leaders from academia, industry and the NIH was held at the NIH on 26 September, 2001 (members invited are listed in Appendix A). This initial meeting generated several recommendations for future consideration that included:

1. *Establish a general biostatistics training grant program*: The future well-trained biostatistician will direct, collaborate on and consult on research in a wide variety of research areas including disease-specific areas, genetics, organism biology, cellular biology, molecular biology, proteomics and neuroscience. It is a strength of the profession that biostatistical methods can be applied in a wide variety of disease areas and research settings. While biostatistics training grant support by individual institutes has been and will remain very valuable, program review includes evaluation of student exposure to research relevant to the sponsoring institute and of the number of graduated trainees who work in the disease specialty. A new training grant program with additional training positions is needed where proposal review is

focused primarily on the quality of the training in biostatistics and the professional success of graduated trainees in academia, government and industry.

2. *Broaden the scope of new or existing training grant programs*: Bioinformatics is a new and rapidly emerging field, commonly based on training in computer science and/or biology. However, genomics has moved the fields of biostatistics and bioinformatics closer together. For example, many of the algorithms utilized in bioinformatics have their origins or theoretical basis in statistics. Biostatisticians are now deeply involved in genomic research and use the computational methods developed for bioinformatics. Existing biostatistics training grants should be broadened to include computer science, mathematics or statistics students who wish to pursue a bioinformatics research career. The shortage of well-trained bioinformaticians is as serious as that in biostatistics. Other areas such as computational biology and neuroscience are also rapidly growing and will require in depth multidisciplinary training.

3. *Increase industry participation*: The biotechnology, pharmaceutical and medical device industries have become major employers of biostatisticians and bioinformaticians, yet the relative support of training of biostatisticians by industry is far less than that of the NIH. To achieve the goal of substantially closing the supply/demand gap in a reasonable time period, innovative methods for industry to participate in training programs are needed at all levels. In addition to providing tuition/stipend funds for the training of biostatisticians, industry funds could be used to support summer internship programs, with a student rotating through one or two industries after having spent one or two summers in an academic program. Industry staff could more actively participate in seminars and workshops (an already-occurring increase).

4. *Improve the peer review system for biostatistics training grants*: Many current training grants in biostatistics are reviewed by a study section that has no or only one biostatistician. Training grants in biostatistics should be reviewed by a panel of peers. Ideally, these biostatisticians should have experience in academic departments with strong research and training programs.

5. *Stimulate undergraduate interest in biostatistics/bioinformatics*: Most undergraduate training programs in mathematics and computer science do not emphasize statistics, biostatistics or bioinformatics. Often, faculty are not aware of the opportunities in these fields, and students lose interest in the mathematics sciences before they have the opportunity to experience and become excited about these fields. Summer workshops need to be competitively supported for undergraduate faculty, either regionally or nationally, to spend time with biostatistics and bioinformatics faculty and become acquainted with the field and be better prepared to stimulate undergraduate student interest. Summer internships should also be supported for undergraduate students to expose them to biostatistics and bioinformatics. During these internships, students might participate in both lectures and collaborative research. The goal would be to stimulate a high fraction of these summer interns to apply to graduate school in statistics or biostatistics, either at the school where they attended the intern program or some other similar department. Other activities to increase the flow of students into graduate programs are described in a recent issue of *Amstat News* [9]. The national pool of students would be enhanced and the much needed supply of domestic students would increase. Since that 2001 meeting, three biostatistics/statistics departments (Boston University, North Carolina State University, University of Wisconsin-Madison) are NIH-funded (by the National Heart Lung and Blood Institute) to host summer internship programs for outstanding undergraduates with quantitative skills and interests in biological science. Collectively, nearly 70 undergraduates have been awarded summer internships each year since 2004. The first co-

hort of undergraduate summer interns are just now applying to graduate programs, indicating that this program in fact is producing new interest in biostatistical careers.

6. *Develop new mechanisms of support for non-residents*: For the immediate future, the shortage of both biostatistics and bioinformatics professionals will not be met realistically by the number of domestic students applying to graduate school in these fields. Most biostatistics/statistics training programs have a high proportion of non-domestic or foreign students, many of whom remain in the U.S. after completing their degrees. Given the enormous and often disproportionate health problems borne by developing countries, and the need to develop quantitative expertise in these countries, training grant programs similar to that implemented by Fogarty Center for HIV should be developed for biostatistics training. Innovative support mechanisms such as supplementing R01 research grants or program project grants could provide much needed short-term training support.

NIGMS WORKSHOP, 5 DECEMBER, 2003

As a follow up to the September 2001 meeting, the National Institutes of General Medical Sciences (NIGMS) organized a workshop on 5 December, 2003 to discuss a novel training program that would address several of the recommendations of the earlier *ad hoc* group. Leaders from academia who attended are listed in Appendix B. This group made additional recommendations to the NIGMS addressing some of the earlier concerns.

The workshop recommended that a new NIGMS training grant initiative was necessary in order to respond to the growing need for a new generation of biostatisticians who are trained in the basic core of biostatistical theory and methods as well as clinical trials and observational studies but also in new and evolving methodologies related to bioinformatics, statistical genetics, genomics, proteomics and beyond. The NIH, and NIGMS in particular, also needed to recognize that new approaches for supporting training may be necessary to be successful. Relying only on current funding mechanisms may not be the best solution.

This training program must by necessity be interdisciplinary, connecting traditional biostatistics training to the biomedical basic, translational and clinical researchers. Many trainees should learn basic concepts of genetics and molecular biology in order to be able to understand the research problem and to make meaningful collaborations. This new training program in biostatistics with an emphasis in 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 real problems faced by these researchers. A major goal is to train students to become independent researchers, advancing the field of statistical research and its application to biomedical research, both basic and clinical, and to be intellectual collaborators who are intimately involved in the research design and data analysis of these biomedical projects. In order to fulfill the need for researchers in this field, another major goal would be to excite quantitatively talented and scientifically curious undergraduates. Some of the key features of such a new training program are outlined below.

A PROPOSAL FOR AN NIGMS TRAINING PROGRAM IN BIOSTATISTICS

In order to respond to existing and growing needs for biostatistics to advance new knowledge about disease etiology, prevention and treatment, it is time for a renewed emphasis in training

a new generation of biostatisticians. Students must continue to be trained in traditional cores of statistical theory and methods, and have some exposure to the traditional applications in clinical trials and epidemiology. However, new courses will need to be developed and some curriculum revision is likely. We encourage innovative plans from institutions that are designed to meet the needs of this next generation of training programs.

While the training program may be administered by a Department of Statistics or Biostatistics, the program must be interdisciplinary. We imagine that in many cases a student may have a mentor or co-mentor beyond the host department, and direct research interactions beyond that department may be quite common. For example, a biostatistics faculty member might be the primary mentor or advisor for the didactic training and dissertation but a biomedical researcher could serve as a co-advisor to help frame the biomedical question and assess whether the methodological development provides a useful scientific advance.

These novel training programs should be able to be completed within 6 years time beyond the bachelor's degree. This training typically would consist of 3 years of formal course work followed by 2–3 years focused on dissertation research. This formal course work training is essential for otherwise students will not have the fundamental theory and methods of their profession. In fact, training programs will have to be selective in their required course work in order to complete this basic training within 3 years after a bachelor's degree for there is an ever increasing amount of theory and methods that could justifiably be added. To not require a rigorous background would result in a very weak and poorly prepared trainee for the methodological challenges that lay ahead.

However, it is also essential that early on, perhaps in their first year, trainees begin to rotate through one or more biomedical research laboratories to gain experience, insight and motivation for the research process, ultimately identifying a stimulating project which would lead to a dissertation topic based on a real biomedical problem. The goal should be for students to gain first hand experience with front line biology research and the applications of biostatistics.

Students eligible for this training program would typically be students in statistics or biostatistics programs but could also include exceptional students from other graduate programs who have special and substantial quantitative skills. These latter students would either have to have the quantitative training already or obtain the training to be able to complete the curriculum as outlined below.

The training grant would typically support students during their first 3 years of required didactic course work in statistical theory and methods. The expectation is that the training advisor or co-advisor would seek funding for the dissertation period from their own statistical methodology grants, other collaborative research grants, institutional funds or perhaps industrial sources.

The ultimate success of the training program would be the success of graduates in academic, federal or private research institutions. Training programs must follow their graduates and evaluate the impact of their graduates on their research environment.

Finally, it is critical that the NIH establish a peer review system for these biostatistics training programs. Currently, many disease-based training programs in biostatistics are reviewed along with many other biological training programs. There is typically, at most, one biostatistician on the review if any at all. The reviews of biostatistics training grants often suffer from a lack of understanding of how biostatisticians are trained or function in their careers. Biologists are used to a year of course work followed by a few years in the laboratory working on their thesis research and then another few years of postdoctoral training. Biostatisticians rarely complete their course work and PhD exams before 3 years but typically spend 2 years on a thesis and then often go directly into a professional position. Thus, the review process should involve faculty who are part of or are familiar with leading biostatistics training programs across the country.

Key elements of the training program are briefly outlined as follows:

1. *Basic theory and methods in biostatistics*
   Since many biomedical problems will likely require the development of new statistical methods, students should be able to read critically the theoretical and methodology literature in statistics and develop new methods as appropriate for the problems that are encountered. 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. Traditional courses might include for example mathematical statistics including theory of estimation and inference, probability theory, linear models including regression, analysis of variance and generalized linear models, non parametric methods and data analysis. Some additional training in clinical trials and epidemiology would be desirable if time in the curriculum can be identified. Typically, this rigorous training would take 3 years to complete after completing a bachelor's degree.

2. *Knowledge of a biology specialty area*
   Since the goal is to train students to not only be excellent collaborators but ultimately be biomedical researchers on their own, students must gain some fundamental knowledge of a biological specialty area. This knowledge can be gained through formal course work and/or through hands on laboratory experience or a combination. For example, for a student emphasizing statistical genetics, formal course work might include courses in statistical genetics, human genetics, and population genetics, computer science courses such as database structures, and bioinformatics.

3. *Collaborator and researcher*
   Whether through course work or through hands on experience, the biostatistician should become part of the research team, contributing at the very initial stages of the research project. Too often the biostatistician is not engaged until the data have already been collected and fundamental design flaws have rendered the data not analysable in any meaningful way. With experience, the biostatistician ultimately might be suggesting research to be conducted in the laboratory by laboratory or clinical colleagues.

4. *Communication and leadership skills*
   Research is increasingly being conducted in teams of individuals with different expertise and backgrounds. It is critical that the next generation of biostatisticians not only have the technical expertise and biomedical knowledge as outlined above but also the skills to communicate effectively with their research colleagues about both statistical and biological concepts. This skill has not been sufficiently emphasized in the past with the result that biostatisticians have not always been as effective as they might have been, or have missed opportunities entirely. Thus, the training program needs to develop methods that can assure students will have the necessary verbal and written skills to be effective in a team research environment. This might include frequent oral presentations at laboratory meetings or department journal clubs and seminars, drafting review papers on selected topics, preparing mock grant applications or attending formal institutional workshops or courses on effective communication. Students must develop a broad range of research survival skills to enhance their chances of being effective and successful.
   In addition, biostatisticians need to have better training in assuming leadership roles in research teams. They will often need to lead and develop statistical research teams and collaborators.

They may also be a principal investigator on a biological research grant, given the major role that biostatistics is playing in all of biomedical research.

5. *Training in research ethics and information privacy*

All NIH training programs now require students to be exposed to research ethics training including topics such as accurate reporting of experimental results, responsible authorship and appropriate attribution of previous results. Recently, the Health Insurance Privacy and Accountability Act (HIPAA) introduced the need for more security, and layers of security for patient data to be used by both the clinical and basic researcher. Since the biostatistician is likely to engage in both areas, they must be very familiar with federal and state laws regarding data privacy and security.

CONCLUSIONS

While most current training programs in biostatistics successfully accomplish the first element of this outline, and may achieve some success in limited or focused training in a specific disease or biological area, few if any formally require all of these elements as essential components of a training program.

With NIH and other federal leadership and with participation by academia and industry, important, near-term progress can be made. All parties have a stake in the need to increase the number of biostatisticians and broaden their training to include strong emphasis in one or more areas of biomedical application. Real progress will require renewed commitment and effort targeted at existing academic biostatistics training programs, undergraduate programs in the mathematical sciences and perhaps even stimulating interest in high school students. New approaches by NIH may be needed. The current challenges are most definitely opportunities. Biostatistics has an unprecedented opportunity to make major contributions to biomedical, environmental and other applied science. The profession must seize the opportunity so that the next generation of excellent biostatisticians is available to meet the science and policy challenges. While the NIH should not and cannot bear the entire burden, as it did four decades ago, it can provide critical financial and leadership support to launch this novel initiative.

POSTSCRIPT

Since the workshop of 2003, the NIGMS has awarded six new training grants for the next generation of biostatisticians and expects to fund a few more. The requirements for these training grants follow in the spirit of the recommendations of the 2003 workshop. While this is a very promising and encouraging development, the volume of new biostatisticians generated by these new awards is still likely to fall far short of the demand. Yet these grants serve as a model for additional training grants.

APPENDIX A: MEMBERS INVITED TO THE 2001 WORKSHOP ON TRAINING BIOSTATISTICIANS

Ed Davis, UNC—Chapel Hill

David DeMets, University of Wisconsin

Demetrius Albanes, NCI, NIH

Rebecca DerSimonian, NIAID, NIH

Scott Emerson, University of Washington
Vicki Hertzberg, Emory University
Steve Lagakos, Harvard School of Public Health
Dick Landis, University of Pennsylvania
Tom Louis, Rand Corporation
Jeremy Taylor, University of Michigan
Barbara Tilley, Medical University of South Carolina
George Williams, Merck

Dennis Dixon, NIAID, NIH
Larry Friedman, NHLBI
Mitch Gail, NCI, NIH
Nancy Geller, NHLBI, NIH
Misrak Gezmu, NIAID, NIH
Rick Mowery, NEI, NIH
Wally Schaffer, OD, NIH
Belinda Seto, OD, NIH
Ram Tiwari, NCI, NIH
Judith Vaitukaitis, NCRR, NIH

APPENDIX B: MEMBERS OF 2003 WORKSHOP ON TRAINING BIOSTATISTICIANS

David L DeMets (co-chair)
University of Wisconsin

Jonas Ellenberg
Westat

Gary Stormo (co-chair)
Washington University

Avner Friedman
Ohio State University

Michael Boehnke
University of Michigan

Toni Kazic
University of Missouri

Emery Brown
MIT Harvard University

George W. Reed
University of Massachusetts

Raymond Carroll
Texas A & M

Richard Simon
National Cancer Institute
National Institutes of Health

George Casella
University of Florida

Eric Sobel
UCLA

Richard Chappell
University of Wisconsin

Scott Zeger
Johns Hopkins University

Marie Davidian
North Carolina State University

REFERENCES

1. DeMets DL, Woolson R, Brooks C, Qu R. Where the jobs are: A study of *Amstat News* job advertisements. *American Statistician* 1998; **52**(4):303–307.
2. Ashikaga T. Salary survey results. *Amstat News* 2002; **297**:10.

3. Ashikaga T. Salary survey results. *Amstat News* 2006; **344**:10.
4. Shettle CI, Gaddy C. The labor market for statisticians and other scientists. *National Science Foundation Report*, Arlington VA: National Science Foundation, 1996.
5. Schaeffer RL. In a world of data, statisticians count. *Amstat News* 2001; **291**:2–3.
6. Sanderson AR, Bernard LD, Thomas BH, Sharon LH. *Doctorate recipients from United States Universities, Summary 1999*. National Opinion Research Centers at the University of Chicago, 2000.
7. *Doctorate Recipients from US Universities*: *Summary Report 2001*. NORC at the University of Chicago. http://www.norc.uchicago.edu/issues/sed-2001.pdf
8. DeMets DL, Anbar D, Fairweather W, Louis TA, O'Neill RT. Training the next generation of biostatisticians. *The American Statistician* 1994; **48**(4):280–284.
9. Iman RL. How to recruit students for statistics programs. *Amstat News* 2001; **291**:65–68.