Public Health Aspects of Bioinformatics and Medical Informatics

By
Abdelmonem A. Afifi, Ph.D.
Department of Biostatistics
UCLA School of Public Health

I was asked by Dr. Arnold Goodman to make a few remarks about the public health aspects of bioinformatics and medical informatics, as part of my introduction to this session. In preparation, I began by searching on line the most obvious journal in each of these fields for appropriate key words. I started with the *American Journal of Public Health (AJPH)*. My search for "bioinformatics" and "informatics" from July 1997 to April 2001 produced a single result, in which the word informatics was used only in the affiliation of one of the authors. I then turned to the journal B*ioinformatics* and searched for the phrase "public health" from January 1985 to April 2001. Six articles turned up, but not one of them addressed the theme on hand. Similar searches were also fruitless. I thus began to wonder, as Rudyard Kipling might have put it, when shall the twain ever meet?

Well, it seems that some meetings have indeed been taking place. For example, the National Library of Medicine has an entry in its Current Bibliographies, entitled Public Health Informatics, which contains 441 citations. The URL for it is http://www.nlm.nih.gov/pubs/cbm/phi2001.html.

A scan of some of the articles in that site makes it clear that this is a nascent area, typical of what happens at the birth of a new field. One way to enter this area is to look for practical definitions of terms. The term bioinformatics is still very much in the process of being defined. Right now it means different things to different people--from processing and storing data produced by sequencing machines, to modeling protein structure and finding functional relationships among new genes. One definition that seems to have gained some acceptance is the one used by the UCLA Bioinformatics Interdisciplinary Program. It reads: "Bioinformatics is the study of the inherent structure of biological information and biological systems. It brings together the avalanche of systematic biological data (e.g. genomes) with the analytic theory and practical tools of mathematics and computer science." This seems to be a sort of a catchall definition. The term "medical informatics" seems to be even more of a catchall phrase. It is understood by the medical community to encompass all of the electronic resources and information available to the health care professional for the purposes of health care, education and research. For example, Associate Dean Neil Parker of the UCLA School of Medicine, in a presentation available on the Web, describes possible scenarios of medical informatics in future clinical practice as follows:

"BetterHealth@Here.Now - Health

- Practitioners will look up Best Practices on-line
- Hospital Infosystems will be available 24x7 through the Internet
- Clinicians will receive new research information directly relevant to their practice
- Physicians will routinely use Computer facilitated diagnostic & therapeutic algorithms
- Physicians will manage similar patient problems using computer facilitated tools "

(http://www.medsch.ucla.edu/som/admin/idtu/Symposium/retrospective/presentations/Neil/sld001.htm).

The emerging fields of bioinformatics and medical informatics are ripe for exploration regarding their public health-related aspects. We need to examine closely some of their potential benefits to the public's health, as well as some of the pitfalls that must be avoided. Two recent articles discuss the role of informatics in public health*. Both emphasize that the focus of public health informatics should be on prevention and on the health of populations. The guidelines regarding cholesterol, recently issued by the National Heart, Lung and Blood Institute (NHLBI), present a good case in point. The guidelines are essentially an algorithm that takes into account not only the value of total cholesterol, but also, LDL, HDL and a host of risk factors such as age, individual and family disease history, as well as behavioral practices such as smoking. Depending on the individual's profile of all these variables, the algorithm issues a recommendation to the patient regarding whether or not taking a cholesterol-reducing medication is advisable. The NHLBI created a web site to assist the public in understanding this complex topic:

(http://www.nhlbi.nih.gov/guidelines/cholesterol/pat_pub.htm).

The user can retrieve information about the role of cholesterol in the body as well as obtain advice on treatment and prevention of heart disease. In one available site, the individual supplies the above profile and the program computes the risk of developing heart disease over the next 10 years.

This is an example of the potential of bioinformatics and medical informatics in disease prevention, one of the fundamental functions of public health. It is within our grasp to be able to generalize this example many-fold. Based on the individual's profile, including such factors as the ones just mentioned, as well as the genetic information that is at the heart of bioinformatics, it will be possible in the near future to formulate guidelines that are tailor-made to each person to help her or him live a healthier life. The guidelines would incorporate conclusions dawn from genetic investigations and population studies, such as the Framingham Study. They would include, among other things, recommendations regarding diet, food supplements, exercise regimens, medications and even choices of occupational and recreational activities. On the curative side, the promise is there for designing treatment protocols that are suited to the specific characteristics of the individual patient. From a public health point of view, we need to begin planning for preventive and curative health services systems that can facilitate this new thinking. In short, bioinformatics and medical informatics can help us come closer to achieving our public health mission, namely, "assuring the conditions in which people can be healthy".

On the other hand, there are a number of challenges that may have to be faced, both in the theoretical and practical domains. In terms of data analysis, the flood of information from genomics, proteomics, and microarrays can overwhelm the current methodology of Biostatistics. For example, one can imagine situations in which discriminant functions are possible to derive, which do an excellent job of discrimination in terms of the training sample, but do a very poor job on future samples. Such prospects present a challenge for the biostatisticians to adjust current methodology to the demands of the oncoming flood of data.

Exploratory methods such "data mining" and other computer-based approaches are tempting avenues for dealing with these data. But such methods raise fundamental philosophical issues that require us to reconsider our statistical paradigm of the Neyman Pearson hypothesis testing methodology based on Karl Popper's philosophy of science. The question is: should we go on hunting expeditions that attempt to induce conclusions from data, or should we, ahead of time, formulate a theory, along with its implied hypotheses, and then proceed with attempts to falsify these hypotheses in an effort to further develop the theory? In an editorial in the journal *BioEssays*, John F. Allen maintains that data mining is doomed from the start because there are no "induction machines" to automatically produce conclusions from data (Volume 23, 2001, pp. 104 – 107). He argues vehemently in favor of continuing to conduct our scientific endeavors on the basis of Popper's philosophy of testable hypotheses. I personally think that this is an extreme position. In fact, an intriguing question is: Can these two philosophies coexist, and if so how? One reasonable compromise that is currently employed by some scientists is to use the exploratory techniques to uncover promising genes and genetic relationships that can later be examined via traditional statistical methods. Such a process takes place routinely in the physical sciences where theoreticians and experimentalists join forces whenever they tackle new terrain.

Finally, as public health professionals, we must address several ethical issues stemming from the future availability of individual biological profiles made possible by bioinformatics and medical informatics. A case can be made that some discrimination based on whether a person smokes or is overweight takes place right now. In addition, it is not difficult to anticipate that the eligibility of individuals for health and life insurance can become threatened by whether they fit certain criteria based on such profiles. An obvious example is that health or life insurance can be denied to people with certain genes that make them more vulnerable to particular diseases. The danger also exists for denying employment based on the person's bioinformatic profile. Thus, in addition to outlawing discrimination based on race, creed and place of origin, our society needs to formulate measures to protect individuals from discrimination based on bioinformatics.

In summary, there is a real potential for bioinformatics and medical informatics to measurably improve the public's health. There are also interesting challenges and dangerous pitfalls that need to be addressed. I believe that now is the right time for scientists, policy makers and public health professionals to a have a serious dialogue regarding these issues. One step in that direction is the newly formed Special Interest Group in Public Health Informatics of the American Medical Informatics Association. The members of that Special Interest Group have already begun this dialogue. Let's all join the fray.

*Yasnoff, W. A., O'Carroll, P. W. et al. "Public Health Informatics: Improving and Transforming Public Health in the Information Age," *J Public Health Practice,* 2000, 6(6), pp 67 – 75.

*Luck, J. "Application of Information Systems to Public Health," in: *Oxford Textbook of Public Health*, Holland, W., Detels, R. and Knox, G., editors, 2001 (Forthcoming), Oxford: Oxford University Press.