Very Large Databases and Web-Based Healthcare Records

Is there an opportunity to advance patient care and health-based research?

BY DAVID A. WALLACE, MD

With the rapid expansion of diagnostic and therapeutic capabilities in our specialty, we ophthalmologists find ourselves increasingly challenged to stay on top of our games. The information we use to make clinical and surgical decisions is fragmented and comes at us from different directions, each with its own bias. Another factor contributing to the current state of clinical medicine is that our medical data systems are also fragmented and incompatible. If we desire to ascertain, for example, the incidence of endophthalmitis after temporal clear corneal cataract surgery in a large population, there is no database to which we can turn to query records. Our individual databases are not compatible or interactive enough at this time to patchwork together such a system.

I am unaware of any database within any medical specialty that is yet able to aggregate clinical or diagnostic information on a very large scale using a Web-based platform. However, the opportunities for advancement of knowledge, identification of cause-and-effect relationships, and improvement in understanding of population statistics regarding discreet health risks is enormous. Think of the possibilities if we create a national (or larger) system to facilitate sharing such information among all practitioners, students, and researchers within our specialty.

Processing large amounts of information used to be difficult and time consuming, as was obtaining access to useful data. Although many medical databases currently exist in some form or another, several more could be created to help us in making healthcare decisions for our patients and to aid our study of new medical technology.

> "It may be reasonable to recognize that the success of an information-dependent entity in the modern world is proportional to the size of its database."

EXAMPLES OF DILEMMAS DUE TO A LACK OF LARGE DATABASES

Clinical Problem

A patient had LASIK 8 years ago and now needs cataract surgery. Optimally, you would like to have the pre-LASIK refraction, topography, and other information to assist with IOL calculations, but the chart at the previous provider's office (or maybe even your own) has been purged or is not available. Would it not be nice if such information were archived online in some format readily accessible by a Web-based interface?

Standard-of-Care Issue

What is the risk of corneal ectasia in the natural population, and what is the additional risk in patients with topo-

graphic evidence of forme fruste keratoconus? Individual practitioners, even those with large practices, may have inadequate data to accurately calculate these risks. As a result, practitioners may make false conclusions about the appropriateness of surgical intervention, and this may lead to imperfect or misguided notions about standard of care in a community, with consequent legal ramifications. Would not it be helpful if there were a national collective database of ophthalmology-relevant data from which one could derive this information?

EXAMPLES OF LARGE DATABASES

- **State motor vehicle departments.** These track drivers licenses, vehicle registrations, and VIN numbers for millions of people and vehicles.
- **The National Library of Medicine's PubMed literature citation database (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed).** This indexes thousands of journals and millions of articles. The database goes back many years and is cross-indexed by author, keyword, specialty, etc.
- **Online music libraries.** Those such as Apple's iTunes (http://www.iTunes.com) warehouse millions of files of songs by tens of thousands of performers going back to the earliest days of recorded music, so that anyone, anywhere can, for $0.99, download any tune at any time.
- **Real estate.** Go to http://www.zillow.com and type in your home address for a description of your property and an estimated current value. This applies to residential properties in all 50 states and reflects a database of more than 60 million individual listings.
- **Major Internet search engines.** Those such as Google and Yahoo index every Web site on the Internet. That's hundreds of millions of sites, some with thousands of pages, and including associated keywords and internal and external links.
- **Google Earth.** Now, for literally any location on the entire planet, Google will give you a satellite image.
- **National Health Insurance Database.** Health insurance carriers warehouse massive amounts of protected healthcare information (PHI). All carriers maintain databases that track every insured (past and present), every encounter with every registered provider, every CPT and ICD code ever submitted, and of course every penny billed and paid. For the major carriers like Blue Cross, Blue Shield and the regional Medicare intermediaries, that's tons of data.

WHAT WORKS AND WHAT DOES NOT

The current paradigm for medical record archiving dictates that each practitioner, group, or institution should maintain and mine its own patient database with cost-intense systems that are not cross-compatible or interactive. As information technology evolves it becomes clear how limiting this approach may really be. If instead patient information were stored in a Web-based national system, the per-provider costs would plummet, and the possibilities for pooled-data analysis or research would become enormous. Of course, this idea describes a new and different paradigm, one that will not appear overnight. But, we should consider some of the large database systems of information that currently exist and contemplate what is possible within the medical arena.

It may be reasonable to recognize that the success of an information-dependent entity in the modern world is proportional to the size of its database. The list of existing large databases intentionally excludes those maintained by nations and their respective police, military, and intelligence agencies, because that information is classified. The size, level of detail, and complexity of those databases can only be the subject of speculation, but surely they are even larger than the examples cited.

WEB-BASED NETWORKS IN OPHTHALMOLOGY

There are already a few examples within our profession of Web-based networks for specific channels of information. Guy Kezirian, MD, has developed refractive surgery outcomes software (Refractive Surgery Consultant or RSC, Scottsdale, AZ) for which a Web-based version exists to serve the interests of Wavelight laser (Wavelight Technologie AG, Erlangen, Germany) users. "RSC Lite" allows users to input data and derive nomogram recommendations from a pooled database of thousands of treatments on similar lasers. My patients and I have benefited from my use of this resource. Similarly, surgeons desiring to become certified with the Acrysof Restor IOL (Alcon Laboratories, Inc., Fort Worth, TX) must register and enter treatment data for their initial 30 cases using a different lens, the monofocal SN60WF lens (Alcon Laboratories, Inc.) using a Web-based interface, available at http://www.previzetech.com.

Pooling and sharing data have intrinsic collective value. This value is recognized in systems ranging from several cell neural networks to the Internet to international anti-terrorist task forces. If anything, we now recognize that the worth of any successful network is far greater than the worth of the component parts disassembled.

Until recently, two factors presented barriers to ophthalmologists' even thinking in terms of such large databases. The first was that the number of data objects tracked, along with the associated dependent and independent variables, is huge, beyond what a typical person can grasp. The second was economics, which were prohibitive until

recently. Thanks to Moore's Law (computer processing speed doubles roughly every year) and its corollary as it applies to media storage (capacity increases geometrically while cost decreases), these barriers are now surmountable.

LET US RUN THE NUMBERS

It seems to me that we have reached an inflection point of sorts where information technology capabilities have exceeded necessary capacity, and the cost has dropped well within the range of affordability. How much storage space would be required to hold, for example, every topographic study for every patient seen in every eye care provider's office in the US in a year's time? Each data file requires about 1 megabyte or $1x10^6$ bytes. Let us estimate on the high side and assume that 4 million studies ($4x10^6$) are performed annually in the US. In order not to underestimate, let us multiply by a factor of 10 just to be safe. The total is $4x10^{13}$ bytes, which seems at first incomprehensible. A megabyte is 10^6 bytes, a gigabyte is 10^9 bytes, and we now have commercial systems that can store terabytes, which are 10^{12} bytes. So, the entire universe of topography, times an "underestimated" factor of 10, is 40 terabytes. The cost in 2006 for a terabyte of disk space is approximately $500. Therefore, at retail rates (which would be overpaying even by current standards and pricing) the cost for 40 terabytes of storage is $20K. It is affordable, or nearly so, and Moore's law (or the storage corollary) will guarantee that the costs come down over time, not stay flat or rise.

WHAT IS POSSIBLE

Medicine is in many ways behind the curve when it comes to implementating large database systems. The following objectives are now at least theoretically possible, from the perspective of cost and information infrastructure:

- archived cradle-to-grave healthcare information for every individual seen or treated in the country;
- better statistical data for risk analysis;
- refinements in clinical medicine through, for example, Web-based platforms guiding the development of excimer laser nomograms and case-by-case surgical planning;
- better methods to identify statistical outliers (eg, eyes with borderline topography, surgeons with unacceptable surgical morbidity, etc);
- better assessment of actuarial risk for large populations (fairer healthcare insurance premiums, possibly fairer or lower malpractice premiums, etc); and
- better accountability systems for providers and payors of healthcare.

If we think that better information allows improved decisions and planning, we should generally be in favor of compiling such large healthcare databases. Clearly, these large systems will not appear overnight, but, if there is enough interest, the technology is within reach and is now economically feasible.

ELECTRONIC MEDICAL RECORDS

Current electronic medical record systems are a hodgepodge of products with inherent incompatibilities and significant limitations that make implementation time and cost intensive. The major players in the database space (eg, Microsoft Corp, [Redmond, WA], Oracle Corp. [Redwood Shores, CA], and SAP AG [Walldorf, Germany]) do not even offer systems in this field. Why? Could it be because free-market economics and common sense do not intelligently guide the warehousing of medical information? We do have well-intentioned private foundations like the Health Information and Management Systems Society (http://www.himss.org) developing strategies and white papers, but we also have meddlesome government regulations like HIPAA that try to shape what we do and how we do it.

There are legitimate concerns about the assembly of large databases, including the protection of personal healthcare information, "need-to-know" access, and the potential for fraud as well as for abuse and/or misuse of the system. In addition, one needs to consider and carefully structure healthcare information systems to respect the rights of the individual (patient, provider, etc) while enabling collective learning and benefiting the population as a whole. I am not suggesting that these systems be developed without attention to such important matters, but I think it makes sense to continue the dialogue about these systems and their potential to benefit healthcare.

LIVING IN A MONITORED WORLD

The Orwellian notion of Big Brother continues to be troublesome. This is now 2006, not 1984, and in many ways we live in a highly monitored world already. The challenge is to participate in creating a system (or systems) wherein the benefits of pooling information far outweigh the risks. In that sense, we can think about this in a fashion similar to decisions about other medical or surgical care. After all, healthcare information exists to serve both the needs of the individual and the interests of the collective such as medical specialties, populations, cultures, or nations. ■

David A. Wallace, MD, is Medical Director for LA Sight in Los Angeles. He acknowledged no financial interest in the products or companies mentioned herein. Dr. Wallace may be reached at (310) 828-2020; drdave@la-sight.com.