Section 43

ONCOLOGY AND THE INFORMATION REVOLUTION

159 ONCOLOGY AND THE INFORMATION REVOLUTION

EDWARD P. AMBINDER, MD

HEALTHCARE POLICY AND INFORMATION TECHNOLOGY

As clinical oncologists, we are experiencing unprecedented challenges to our research and therapeutic and practice management skills brought on by the necessity to manage the geometric increase in medical information and the striking changes in the clinical practice of oncology demanded by managed competition. The clinical oncologist must be a clinician, researcher, educator, businessperson, statistician, healthcare administrator, and " informatician," who must interact with the whole patient, both as a specialist and as a primary-care physician. These skills must be assimilated into our daily clinical practice and research activities, while we continue to improve our quality for a much more sophisticated consumer and payer.

Information management has become the most frustrating aspect of our professional life. Cancer patients receive intricate multi-modality therapies for their complex diseases, requiring greater access to, and tracking of, detailed medical data. Disease management clinical guidelines for cancer patients are being introduced which require the micromanagement and economic awareness of patient care.[1]

Managed competition as well as clinical research trials demand more intense recording of medical documentation, while reimbursements and time available for each patient are curtailed. Clinical trials must not only show progression-free and overall survival benefits but now must also display cost-effectiveness and patient satisfaction.[2,3]

The digitalization of information, the ability to network or connect computers, and the rapid electronic interchange of information on a worldwide basis are recognized as the hallmark of a new order in society. For oncologists to master these profound changes, it is necessary to comprehend the medical tools of the Information Age. These tools include the computer, which acts as the information gateway and integrator, the electronic medical record, which is the translator and repository of our clinical information gathering, and the Internet's World Wide Web (WWW), which coordinates interactive information resources and personal communication. Information technology and personal computers have transformed every other profession and are now revolutionizing medicine as well.

Today, health-care reform in the United States requires us to reconcile the needs of 40,000,000 uninsured patients, the individual insured patients and their health-care organization in this cost-containment environment and to apply rapidly expanding molecular and genomic knowledge for an aging and mobile population with rising chronic disease rates. The delivery of patient care in the United States is disjointed, when considering the mobile society, patients who change practitioners and health plans every 3 years on average, the absence of coordination of the patient's longitudinal medical record, and the lip service given to the screening and prevention of disease. Health-care delivery automation, with the exception of practice management, results reporting and financial software, is shamelessly underutilized despite the nonmedical world's rapid acceptance of the tools of the Information Age. Moore's Law first formulated in the 1950s accurately predicted the doubling of microchip technology capabilities and the halving of computer costs every 18 months. Similar changes are occurring in health-care delivery utilizing computerization, electronic medical records, and the Internet. Most exciting is the the recent merging of biomedical and silicon chip technology with the advent of the biochip, where the entire human complement of 80,000 genes will be simultaneously and inexpensively assayed for abnormalities and biologic activity.

If our health-care system is to survive and thrive in a rapidly changing medical universe, all the participants—providers, patients, administrators, researchers, educators, payers, industry and the government—must understand the laws of the Information Age. They must invest the time and energy to harness the vast untapped powers of computers, electronic medical records, and the Internet. Combining the huge cost-saving potential of computerization and the Internet telecommunication technology will provide a logical framework for health-care in the next millennium.

With secured and confidential electronic medical records and the acceptance of community health information networks (CHINs) that electronically link, on a regional basis, all patients, practitioners, payers, hospitals, benefit managers, commercial laboratories, and drug stores, our health-care system will become cost effective and efficient. Most of these networks also provide free home pages, secure VPNs (virtual private networks) for large physician groups, access to continuing medical education and medical journals, and hot links to e-commerce alliances. Among the key questions to be answered are: can the internet solve the major problems of today's electronic data interchange (EDI), such as incompatible technologies and lack of standard formats? How will companies solve data security and privacy problems that are both perceived and real? Will companies' relationships with payers and pharmaceutical benefit managers make physicians reluctant to use their services?

Building on fundamental strengths and with a common spirit of compromise and sharing, we could look forward to a health-care system that maintains its world leadership in the delivery of oncologic health-care, the production of clinical and basic science oncologic research and the oncologic education of health-care providers and patients.

MEDICAL AND ONCOLOGIC INFORMATICS

Medical informatics uses information science and its technologies to deal with the "cognitive, information processing and communication tasks of medical practice, education, and research."[4] Medical oncology, with its strong emphasis on data collection and analysis, is positioned at the forefront of clinical trials requiring quantitative methods to record clinical signs, symptoms, toxicity, and outcome. In its training programs, and carried over into clinical practice by its trainees, there is a heavy emphasis on research protocols requiring a time-oriented tabular flowsheet format that is ideal for computerization. Increasing documentation for cooperative groups, government bodies, pharmaceutical houses, and institutional review boards demands computerized records. Health-care costs have escalated dramatically, particularly for cancer patients, who require complex diagnostic and therapeutic procedures, making them a prime target for cost-containment efforts by third-party payers and the federal government. Today, the computer is our most efficient tool for obtaining timely information and has become our most effective decision-making tool. The computer has become an integrated information device that has merged printers, scanners, facsimile machines (faxes), copiers, telephones, video and audio systems, and filing cabinets. Interested readers are referred to *Medical Informatics: Computer Applications in Health Care* for a comprehensive review of this field.[5]

The first wave of computers increased productivity by giving users new tools, but the sheer mass of data actually tended to decrease productivity by generating data faster than the user could analyze it. Early users had to overcome the terror of learning a whole new language and approach. They preferred to delegate computer tasks to assistants or to their computer-literate colleagues. The business community was the first group to recognize the importance of using computers to manage information. It was not until recently that the health-care industry and providers began to utilize computers for information analysis and management. As oncologists become more comfortable with computers and the electronic distribution of computer-based information and as patients take more responsibility for their medical decisions, health care, as noted by Kassirer, will utilize online computer-assisted communication among patients, medical databases, and physicians to replace a considerable part of the care now given in person.[6] Today's machines address the handling of data analysis by using intuitive interfaces. In the future, computers more nearly will simulate the "whole person paradigm," whereby data can be input by voice, pointing, writing, or gestures, and different languages can be freely translated.

Computers will never replace the medical oncologist. What a computer can do is take a repetitive, time-consuming, and exhausting task and do it over and over again, perfectly, without getting tired or bored, freeing the oncologist to think and reason about medical problems. The computer revolution will enable the oncologist to work efficiently with up-to-the-minute information that is presented clearly and informatively, and ultimately to manage the cancer patient more effectively.

To understand how the decreasing cost and increasing efficiency of processing, storing, and transmitting information have changed more rapidly than any prior human endeavor, it is necessary to be familiar with the three basic laws that have governed the Information Age. Continuing unabated since 1950, Gordon Moore's Law states that the performance of semiconductor technology, as exemplified by microprocessor chips, where data is stored and manipulated, will double every 18 months. As demonstrated by the Internet, Bob Metcalfe's Law states that the power of a network rises by the square of the number of terminals attached to it. Finally, the Law of the Telecosm, as formulated by George Gilder, combines Metcalfe's Law with Moore's law of the Microcosm.[7] It ordains that the cost-effectiveness of computer terminals will rise by the square of the number of additional transistors integrated on a single chip. Amplified by the Law of the Microcosm, the Law of the Telecosm signifies the rise in the cost-effectiveness of a network in proportion to the resources deployed on it and the number of potential nodes and routers (connectors) available to it. In the 50 years since the development of (ENIAC), the first electronic computer, the efficiency of information technology has increased 100 octillion times.

INFORMATION OVERLOAD

Oncologists are faced with data and information overload, making it next to impossible to locate pertinent data in an efficient manner. De Solla Price estimates that the medical literature doubles every 10 to 15 years, and over 22,000 medical journals and 17,000 medical textbooks are published annually.[8] Thus, knowledge management requires a systematic, digitized, organizational approach.

The amount of information that the brain can register has been estimated at 50 bits per second, which is a fraction of what the sensory organs can process. To prevent drowning in the tidal wave of information coming our way, we will need to improve dramatically the methods we use to screen out unwanted information. This may even be done for us by computers talking to other computers using "knowledge navigators" or "knowbots," which will be computer programs that can find items of interest and banish those not wanted.

OVERVIEW OF COMPUTERS

A computer receives information from the keyboard or mouse (input devices), stores it in its logic board (central processor) or hard disk (storage medium), manipulates the information in its logic board, and delivers it through its video display (output device). For broader information, the reader is referred to an excellent concise overview of computers by Toong and Gupta.[9]

Bits and Bytes The electronic currency of the computer is the electric signal, and it is based on binary mathematics, that is, something is positive or negative, 1 or 0, on or off, yes or no, black or white. In contrast, our numbering system relies on a base of 10 with 10 numbers (1 to 9 and 0) and is called the decimal system. Computers store information in "bits" (binary digits), which are defined as the smallest units of information. A bit can represent one of two states: 1 (on) or 0 (off). A bit can also represent a dot on a monitor screen (pixel), which is either turned on or turned off to represent black or white. Eight bits make a "byte" that can represent any of 256 values since its 8 bits can be on or off in 256 different combinations. Bytes can be used to represent a letter, number, symbol, or a horizontal row of 8 pixels on a monitor as displayed in a paint or drawing program, or used as an aid for counting or summing. Finally, a byte can represent a code to tell the machine what to do, that is, a carriage return or tab. A "word" is defined as a sequence of bits that can be assessed as a unit by the computer. Current computers use a 32- or 64-bit word size that can utilize 4 or 8 bytes of information at a time.

One kilobyte equals 1,024 bytes, or approximately 1,000 characters. A double-spaced typewritten page requires 2,000 characters or 2K. The kilobyte and megabyte have become the standard units of measure for the storage capacity of disk drives, memory capacity of computers, and disk space allotted to files.

Central Processing Unit The central processing unit (CPU) is the conductor of the computer that processes information. Its size determines the computer's speed and the amount of information that it can handle. It takes digital information sent over the data bus (an internal communication line) from the input device, performs basic arithmetic and logic functions, and sends the information back through the data bus to the output device or to the computer memory for storage. The CPU utilizes RAM (random access memory), which operates only when power is applied to the computer and which stores both instructions and data, and ROM (read only memory), which stores permanent instructions for the operation of the computer that are indelibly etched into a chip during manufacture. RAM is the memory that you can read and write to, whereas ROM is like a series of little programs built into

Table 159.1. Input and Output Devices for the Computer

Device	Input Device	Output Device
35 mm slides/projector	*	*
Barcode reader	*	
Cellular phone	*	*
CD-ROM	*	
Digital pen	*	
DVD (digital video device)	*	
Fax	*	*
Graphic pad	*	
Joystick	*	
Keyboard	*	
Keypad	*	
Microphone	*	
MIDI sound editing	*	*
Monitor or video display terminal (VDT)	*	
Mouse	*	
Medical imaging devices	*	*
Other computers	*	*
Overhead projection pads	*	
PDA (personal digital assistant)	*	*
Printer (laser, dot matrix or ink-jet)	*	
Plotters	*	
Radio	*	*
Scanner	*	
Storage device (disk or tape)	*	*
Tape recorder	*	*
Telephone	*	*
Television	*	*
Touchscreen	*	
Trackball (stationary mouse)	*	
Trackpad (stationary mouse)	*	
Video camera	*	
Video recorder	*	*

the computer that you can read from but not write to, and that gives the computer its personality. RAM and ROM are stored on silicon chips. The time required to read from or write to these chips is called the access time and is measured in billionths of a second.

Hardware Output devices deliver information to the user or other computers and include the television screen or monitor, printer, modem (used to interface with the telephone system), film processors, video recorders, or musical instruments. Storage devices, which include floppy disk drives, hard disk drives, tape drives, compact disk–read only memory (CD-ROM), and erasable laser disk readers, keep programs and documents in permanent or transient memory. The user can enter data into the computer by writing, marking, typing, pointing, speaking, or using a direct link. Input devices, which accept information, can take audio, video, photographic, xerographic, radiographic, typographic, telephonic, or fax data and put the data in digital form for the computer's central processor to assimilate. Digital information is then transmitted to output devices that can display the information. These devices are listed in Table 159.1. The computer itself, with its electronic components and the attached input/output and storage devices make up computer hardware.

Software Software constitutes the instructions that tell the computer what to do. It is stored on disk media and is read into RAM memory when needed for a particular application. Software includes the operating system, which coordinates the flow of information between input and output devices, the CPU, memory storage devices, applications (programs written for a specific task, such as word processing), utilities (programs that make the computer work more efficiently), and programming languages (for communication between the computer and the programmer).

Documents are words, numbers, or pictures created by the user. A summary of the common types of software applications of interest to the oncologist is given in Table 159.2.

Computer Memory The typical internal memory of a computer today uses from 64 megabytes (MB) to 1,000 MB double in-line memory module (DIMM) chips that can hold in RAM memory a total of 64 to 1,000 MB of information. A floppy disk holds 1.4 MB of memory. A hard disk can store up to 51 GB (gigabytes = 1,000 MB) costing 3.33 cents per MB. A CD-ROM laser disc can hold 700 MB. A digital video disc (DVD)-ROM can hold 5.5 GB of data, enough to permanently and durably store 3.14 million citations. It has become a practical storage medium for permanently storing up to 2.36 million typewritten pages or 4.7 billion characters.

Application Architecture Operating systems help control how computers handle the flow of data and assist the user-computer interaction. Networking operating systems handle the flow of information between computers and peripheral devices, such as the printer. Desktop operating systems coordinate the flow of information within the individual computer and its component parts and devices that are connected with its user interface. Microsoft Corporation's Windows NT and Sun's Solaris for UNIX and Linux are the major networking operating systems, while Apple Computer Inc.'s Macintosh System and Microsoft Corporation's Windows are the leading desktop operating systems. Network operating systems feature hard-disk storage capability, client (end user) and server (fileserver or main storage location of data), network topologies (like Ethernet or Token Ring protocols), wide area network (WAN) gateways connectivity, remote management tools, security, password protection, back-up tools, forced log-off after specified time, and printing features. Each system requires its own rules that have made the transfer of information and portability of applications between operating systems difficult. Newer systems are beginning to adopt a consistent framework across different machines or different operating systems of the same machine, however, which should simplify the learning curve for using computers. Application architecture is a set of tools for developing applications by controlling data, file, database, and network organizations. It permits the program to interact with the operating system. This involves managing the keyboard, display, disk drives, file systems, printer ports, serial ports, small computer system interface (SCSI) and firewire ports, and memory. As popularized by the Apple Macintosh, current trends are for a consistent graphical user interface (GUI) so that the computer screen will look the same in all programs, and the common conventions for using the computer will be adhered to by program developers. A GUI consists of graphical icons for identifying objects for manipulation, such as a file, a folder, or a trash can, pull-down menus for choosing commands, a scroll bar for navigating through a document, a window which is a square frame where one works within a program, and a mouse, which is used as a pointing device for manipulating the icons, menus, windows, and scroll bars and which has a clicker that causes an action. A single click selects an item or action, double-clicking simultaneously selects and starts an action, and dragging moves items on the screen or selects groups of items. The use of a mouse frees the user from having to memorize the commands to execute all the actions described.

Networks Networks allow one computer to talk to another computer, permitting electronic mail, exchanging files, sharing printers, and accessing the Internet. For an orderly exchange of information to take place, two common languages are used—Ethernet and Token Ring. Ethernet combines "packet switching," which breaks data up into chunks that contain an identifying address, and a single wire that broadcasts the same message to all computer addresses, but only the computer with the correct address will pick out the information intended for it. Currently, Ethernet communicates at 10 million bits per second. Token Ring has two advantages over Ethernet. It can communicate faster (16 million bits per second), and it can minimize traffic tie-ups on the shared line by using a special packet called a *token,* which continuously circulates and causes packets of information to take turns and avoid a collision. Today, because of its lower price and its open standards, Ethernet has become the most popular networking standard, with Token Ring relegated to large organizations that combine mainframes with microcomputers. As telephone-type wire supplanted coaxial cable and hub-and-spoke configurations proliferated, computer users in close proximity connected their computers to local area networks (LANs). For computers that speak different languages,

Table 159.2. Medical Uses of the Computer

Medical Uses of Computers	Examples		Discussion
Communications	Bulletin boards Bibliographic data bases Voice mail	Electronic mail Electronic forums	Communication systems to store information, retrieve selectively and transmit.
Database management system			A data base can be defined as a program that will allow you to store information for rapid and easy retrieval. It consists of elements called fields (pieces of information) records (number of related fields) and files (numbers of records making up a meaningful whole). Each item may then be searched for, sorted and combined according to given criteria.
Medical records	Admission/Discharges Demographic information Insurance eligibility Office records Hospital records Bedside terminals Patient monitoring Patient education/reminders Patient allergies Body surface area calculation Graphic data representation Image analysis Integrated patient care plans Clinical research protocols Census reports	Tumor registry Pain management tools Protocol management Links to hospital databases Scheduling tests Cumulative drug dosing Orders Blood products: history, matching, utilization Flowsheets Laboratory data Patient summaries Specialty encounters Malpractice documentation Legal documentation	
Pharmacology	Drug-drug interactions Patient drug list Drug ordering	Drug monitoring Antibiotic screening	
Clinical decision support	Artificial Intelligence Natural language processing Computer-aided diagnosis	Computer-aided treatment Medical decision making Consultation	Advice system applies information to assist in diagnosis and management.
Laboratory monitoring	Control and data analysis Physiologic monitoring		
Management systems	Practice management Quality assurance	Resource management Utilization review	
Office Management	Electronic banking Electronic billing Accounts payable Accounts receivable Financial analysis	Financial management Inventory Return visit reminders Patient schedules Physician schedules	
Graphics	Painting or Bit-mapped		A "painting" program uses a mosaic of differently colored or monochrome (gray-scale) dots whose pattern makes up the picture. They are useful for paint programs, electronic retouching programs and scanner software.
	Drawing or vector-based or object-based		Drawing refers to a process of creating an image by mathematically mapping the coordinates of geometric shapes in relationship to a fixed coordinate representing height, width, and, in the case of three dimensional objects, depth. They are useful for general purpose drawing programs, computer-aided drafting and design (CAD), three-dimensional modeling and rendering, illustrations, and presentation programs.

Continued on next page

Table 159.2. ***(continued)***

Medical Uses of Computers	Examples		Discussion
Electronic art			Painting or drawing programs create artwork or import them from a scanner or from clip-art programs. Images can be modified in black and white, gray-scale or color.
Image processing	Radiographic images	Pathologic images	Allow modifications and filtering of files brought in by a scanner or dedicated image processing hardware. They can work with 256 gray/color levels. Useful for importing radiographic pictures
Presentation	Slides Handouts Television projection systems	Overhead projection pads	
Scanning	Documents Graphics	Optical character recognition	
Statistics	Analysis of variance Cluster analysis Contingency tables Correlation Data handling Graphics Linear programming Log-linear models Multi-dimensional scaling	Multivariate analysis Nonparametric tests Path/latent variable analysis Regression Reliability/test item analysis Survival analysis T-tests Time-series analysis	
Spreadsheets			A spreadsheet is a ledger sheet whose boxes can be filled with numbers that are linked together by formulas and can be automatically updated when changes in a box is made. Each cell has a value rule specifying how its value is to be determined. Every time a value is changed anywhere on the spreadsheet, all values dependent on it are recomputed instantly and the new values are displayed.
Word processing			A word processor is a program used for creating, editing, formatting, storing and generating final copies of documents. It may have utilities to include specifying typeface, page layout, table generation, spelling, thesaurus, and grammar checkers. It should have the ability to mix fonts and graphics.
Publishing	Electronic textbooks	Electronic journals	
Project Management			
Programming			
Grant proposals			
Handicapped patients	Input and output devices		
Music			
Organizational charts			
Text retrieval or search	Medical literature Medical textbooks	Medical databases	
Educational	Programmed instruction Computer-aided instruction Video simulation Courseware	Patient education Examinations Patient instructions	

which may be situated at vast distances from one another, to communicate, a method to send files and messages was needed. By routing data packets among different types of networks using the Transport Communication Protocol/Internet Protocol (TCP/IP), multiple links among networks were created, and this gave birth to the Internet.

The amount of information that can be transmitted through a wire is determined by the wire's bandwidth. The higher the number of messages or the larger the packet of information (i.e., graphics or video), the slower is the network. Thus, future hospital networks that will carry images of computed tomography (CT) scans, pathology pictures, or video conferences, must have broader bandwidths that can carry data at a rate of 10 million bits per second with fast Ethernet switching today, and 1 to 2.5 billion bits per second using the asynchronous transfer mode (ATM) in the near future. Whereas other networking schemes transfer packets of any size, ATM restricts the packet to 64 bits, thus simplifying the identifying and routing of information. Packets can be any digital medium, such as voice, data, or video. ATM has been adopted by the telephone, cable, and computer companies, so it will become the unifying networking scheme.

PERSONAL DIGITAL ASSISTANT

As exemplified by the Palm pilot computer, personal digital assistants (PDAs) are hand-held devices that combine portability, computer processing power, pen-based interaction, and handwriting recognition. They also include personal organization programs, such as schedulers, calculators, reminder lists, Rolodex files, financial and travel tools, document and information retrieval capability, and wired and wireless communication by phone, modem, fax or beeper. The oncologist will use these instruments to hold personal, financial, and patient medical data. These devices can synchronize data with desktop computers and can use cellular modems for information updating and retrieval over the Internet.

Telecommunications and Computer Networks Telephone lines that recognize analog or acoustic information carry voice signals at the rate of 64,000 bits per second and data at only 300 to 56,000 bits per second. In the future, fiberoptic networks will carry data at speeds up to 1 trillion bits per second. To send the digital information produced by the computer over a telephone line, which requires acoustic or wave signals, a modem (short for "modulator/demodulator") that can dial the telephone and establish the initial connection is used. Thus, the computer can send and receive messages and documents to and from other computer networks using telephone lines. Bits and bytes are transferred from one computer to another over a serial communications line, where the bits in a given byte are sent individually. This line has two cables carrying bits of information identified as the presence or absence of electrical voltage from one computer to a remote computer and back.

Standards for frequencies, codes, and speeds have been established to assure the correct transmission of information between computers. The *baud* rate is the rate at which data are transmitted from one place to another. One baud is 1 bit per second. Since 1 character equals 1 byte or 8 bits, dividing the baud rate by 8 will give the number of characters sent per second. Most modems today transmit at 2,400 to 28,800 baud or 300 to 3,600 characters per second.

For direct digital communication between computers, one uses a LAN to link with other computers within a short distance to permit the sharing of common devices, such as printers or storage devices. For longer distances, a wide area network (WAN) is used, such as for linking the oncologist's private office and the hospital. Connecting LANs and WANs by gateways can link the oncologist's office to any computer anywhere in the world.

A telecommunications program tells the computer how to interact with the modem and how to send and receive information and identify itself to the other computer. These instructions may be set up to be performed automatically and unattended by using a "macro" language. For example, the unattended computer can call a remote computer at a specified time every day, send (upload) mail to an individual's mailbox, check one's electronic mailbox, send the incoming mail to one's computer (download), and terminate the session (log off).

To increase efficiency and reduce cost, telecommunication carriers, such as Tymenet and Telenet, break up computer communications into packets and mix these packets so that multiple conversations can take place over the same line. Thus, one dials these telecommunication carriers via a local phone number and types in the address code for a particular online database and one's password to enter the desired system.

Electronic mail (e-mail) facilitates the transfer of information within work groups. Documents can be enclosed with memos, a memo can be sent simultaneously to one or all members connected to the network, memos can be received and sent at a specific time, voices can be annotated to the memo, and individuals can communicate simultaneously by conferencing. Oncologists participating in writing a protocol can review and annotate the document at their convenience, with the master copy available to all and each individual's contributions recorded.

INTERNET

The Internet, with 130,000,000 users and growing at a rate of 2% per month, is a global, independent but cooperative network of government, university, research laboratory, corporate, and private computers communicating by e-mail, bulletin boards, and discussion groups. It has the capability to permit searching data base files and retrieving information and software programs. It is fast becoming a post office, telephone system, and research library for its users. The reader is referred to an excellent introduction to the Internet for physicians by Glowniak and Bushway[10] and an editorial by Kassirer.[6] Regional services or national services such as America On-Line (AOL) offer complete Internet access.

Usenet is a bulletin-board system that groups together Internet messages that pertain to a given topic, making discussion groups, newsletters, focused scientific groups, and subscription lists a reality. The information can be threaded on a related topic, distributed automatically to specific groups, deleted selectively by the system administrator, and have expiration dates attached.

The WWW uses a client-server networking scheme which allows remote requests for data from any client personal computer to the host computer, which treats the requesting computer as if it were on its own local network. The WWW displays its arrays of text, data, sounds, and video in an application called the *browser*. Netscape's Navigator and Microsoft's Explorer control the browser environments. They utilize a graphical interface to search and retrieve and display information from any Internet-linked source that is linked to the underlined words or images. They can also be used by anyone to publish information on the WWW. It has provided a unified interface to the diverse protocols, data formats, and information archives used on the Internet. It attaches the Internet address of the computer that stores the documents desired, the document's location on that computer and the language which that computer understands to send the information back to the requester. Any information located anywhere in the world with an Internet connection can be interconnected, permitting users to travel through the information by clicking on the hyperlinked words. As of July 1999 the WWW has 2.8 million sites holding 6 trillion characters and 800 million pages of information consisting of 15 terabytes (10^{12}) of data and 180 million images. Of these sites, 83% have commercial information, and 6 % have scientific and educational content.[11]

The Internet is becoming the principal method for communicating between oncologists. Most cancer cooperative groups use it for communication among their investigators. Some oncologists send their patient's computerized medical records with all reports and actual imaging studies to other physicians and hospitals over the Internet. As the capability of sending 1 to 100 gigabytes of information per second on the Internet becomes a reality, digital audio and video equipment will permit live transmission of scientific conferences on the Internet. Virtual meetings with audio and video presentations and actual slides and posters are becoming commonplace.

Electronic publishing over the WWW is becoming commonplace, where abstracts, selected articles, or entire journals are made available on the date of publication. Aside from archiving issues, some journals permit subscribers to do full text searches, set up personal archives for saving articles and search results, and download articles in a printed format identical to the typeset pages. Pubmed, a version of MEDLINE cre-

Table 159.3. World Wide Web Resources for the Oncologist

Oncologic Resources	Addresses (URL)
American Cancer Society	http://www.cancer.org
American Society of Clinical Oncology	http://www.asco.org
ARC information on cancer	http://www.arc.com/cancernet/cancernet.html
ATSDR Cancer Policy Framework	http://atsdr1.atsdr.cdc.gov:8080/cancer.html
Breast Cancer Information	http://nysernet.org/breast/Default.html
Cancer (Medical Specialties)	http://galaxy.einet.net/galaxy/Medicine/Medical-Specialties/Cancer.html Cancer
Genome Anatomy Project	http://www.ncbi.nlm.nih.gov
Cancer Information Network	http://www.cancernetwork.com
CancerGuide: Steve Dunn's Cancer Information Page	http://bcn.boulder.co.us/health/cancer/canguide.html
Detecting Breast Cancer	http://nysernet.org/bcic/detecting/detecting.html
Fred Hutchinson Cancer Research Center	http://www.fhcrc.org/
H. Lee Moffitt Cancer Center	http://daisy.moffitt.usf.edu/
IARC	http://www.iarc.fr/
Journal of Clinical Oncology	http://www.jcojournal.org/
Journal of NCI	http://cancernet.nci.nih.gov/jnci/jncihome.htm
M.D. Anderson Cancer Center	http://rpisun1.mda.uth.tmc.edu
National Cancer Institute CancerNet	http://www.cancernet.nci.nih.gov/
National Comprehensive Cancer Network	http://www.nccn.org
NCCS guide to cancer resources: Cansearch	http://www.cansearch.org/
NCI clinical trials	http://cancertrials.nci.nih.gov
OncoLink, The University of Pennsylvania Cancer Resource	http://cancer.med.upenn.edu
Oncology Online	http://www.mecklerweb.com/onco/home.html
Quick Information about Cancer for Patients and Families	http://asa.ugl.lib.umich.edu/chdocs/cancer/CANCERGUIDE.HTML
Regional Breast Cancer Support Groups	http://nysernet.org:80/breast/nabco/resource-list/support-groups.html
SEER	http://www-seer.ims.nci.nih.gov/Publications/CSR7393/
State of the Art Oncology in Europe	http://telescan.nki.nl/start/
Telematics Services in Cancer	http://telescan.nki.nl/
The Genetics of Cancer	http://www.cancergenetics.org
General Medical Resources	
American Medical Association	http://www.ama-assn.org
Atlas of Hematology	http://pathy.fujita-hu.ac.jp/~ichihasi/Pictures/atoras.html
BioMedNet	http://www.cursci.co.uk/BioMedNet/biomed.html
Center for Disease Control	http://www.cdc.gov/
CenterWatch Clinical Trials	http://www.CenterWatch.com/
Computer-Based Patient Record Institute	http://www.cpri.org/
Diagnostic Test Information Server	http://dgim-www.ucsf.edu/TestSearch.html
Duke University Medical Center	http://www.mc.duke.edu/ftp/standards/html
DXplain differential diagnosis	http://camis.stanford.edu/people/bdetmer/dxplain.html
Emerging Infectious Diseases	http://www.cdc.gov/ncidod/EID/eid.htm
Federally-Funded Research in the U.S.	http://medoc.gdb.org/best/fed-fund.html
GASNet Anesthesiology Home Page	http://gasnet.med.nyu.edu/HomePage.html
Grants (Reference and Interdisciplinary Information)	http://www.einet.net/galaxy/Reference-and-Interdisciplinary-Information/Grants.html
Health network	http://www.sarnoff.com/sarnoff/spin-off/healthnet.shtml
Health Services/Technology Assessment Text (HSTAT)	http://text.nlm.nih.gov/index.html Hospital.Net http://hospital.net/
HospitalWeb	http://dem0nmac.mgh.harvard.edu/hospitalweb.html
HyperDOC: The National Library of Medicine (NLM)	http://www.nlm.nih.gov/
Johns Hopkins Hospital	gopher://welchlink.welch.jhu.edu/
Jonathan Tward's Multimedia Medical Reference Library	http://www.tiac.net/users/jtward/index.html Lancet http://www.the lancet.com Mayo Clinic http://www.mayo.edu/
MCW International Travelers Clinic	http://www.intmed.mcw.edu/travel.html
Medical Center	http://www-sci.lib.uci.edu/~martindale/Medical.html
Medical Matrix- Guide to Internet Medical Resources	http://kuhttp.cc.ukans.edu/cwis/units/medcntr/Lee/HOMEPAGE.HTML
Medicine – TheYellowPages	http://theyellowpages.com/medicine.htm
Medline	http://ncbi.nlm.nih.gov/cgi-bin/medline Medscape http://www.medscape.com/
MedSearch America's Physician Finder Online	http://msa2.medsearch.com/pfo/
Morbidity & Mortality Weekly Report	http://www.crawford.com/cdc/mmwr/mmwr.html
National Center for Health Statistics	http://www.cdc.gov/nchswww/default.htm
National Guidelines Clearing House	http://www.cansearch.org/
National Institutes of Health	http://www.nih.gov/
National Library of Medicine HyperDOC	http://www.nlm.nih.gov/
National Science Foundation	http://www.nsf.gov/
New England Journal of Medicine	http://www.nejm.org/ PDQ http://cancernet.nci.nih.gov/pdq.htm
Pharmaceutical Information Network	http://pharminfo.com/ PharmWeb -
Pharmacy Information Resources	http://sunsite.unc.edu/pwmirror/
Physician's Online	http://www.po.com/Welcome.html
PosterNet(tm)	http://pharminfo.com/poster/

Continued on next page

Table 159.3. ***(continued)***

Oncologic Resources	Addresses (URL)
PubMed	http://www.ncbi.nlm.nih.gov/PubMed/ Renalnet gopher://ns.gamewood.net/11/renalnet
The British Medical Journal	http://www.bmj.com/bmj/
The Mount Sinai School of Medicine	http://www.mssm.edu/
The National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
The Virtual Hospital	http://vh.radiology.uiowa.edu/
The Whole Brain Atlas	http://www.med.harvard.edu:80/AANLIB/home.html
The World-Wide Web Virtual Library: Biosciences	http://golgi.harvard.edu/biopages/all.html
UCI MED-ED	http://orion.oac.uci.edu/~sclancy/med-ed.htm
UCSF Division of General Internal Medicine	http://dgim-www.ucsf.edu/
US Government Healthfinder	http://www.healthfinder.gov/
Biomedical Resources	
Biologist's Control Panel	http://kiwi.imgen.bcm.tmc.edu:8088/bio/bio_home.html
BMEnet Biomedical Engineering Resource	http://fairway.ecn.purdue.edu/bme/
EMBnet: European Molecular Biology Network	http://beta.embnet.unibas.ch/embnet/info.html
Mouse Genome Informatics Project	http://www.informatics.jax.org/
National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
Primer on Molecular Genetics (Department of Energy)	http://www.gdb.org/Dan/DOE/intro.html
	Directories, Indices & Searching Tools
Argus/University of Michigan Clearinghouse	http://www.lib.umich.edu/chhome.html
AT&T Internet Toll Free 800 Directory	http://www.tollfree.att.net/dir800/
Awesome Lists (makulow@trainer.com)	http://www.clark.net/pub/journalism/awesome.html
Colorado Area Research Library	http://www.carl.org/
Directory	https://www6.internet.net/cgi-bin/getNode?node=1&session=-144
E-mail Discussion Groups	http://www.nova.edu/Inter-Links/listserv.html
FileList	http://l0pht.com/~spacerog/filelist.html#JustCoolStuff
Four11 White Page Directory (SLED)	http://www.four11.com
Home Page - TheYellowPages.com(tm)	http://theyellowpages.com/
InfoSeek Home Page	http://www.infoseek.com/
Home Internet Guide by Franklin	http://ug.cs.dal.ca:3400/newbie.html
Internet Search	http://home.netscape.com/home/internet-search.html
Lycos Server- Carnegie-Melon	http://lycos.com Phonebooks gopher://merlot.gdb.org/11/phonebooks Search E-mail
addresses	http://www.four11.com/
Search the World Wide Yellow Pages(tm)	http://www.yellow.com/cgi-bin/SearchWWYP
Stanford Netnews Filtering Service	http://woodstock.stanford.edu:2000/
THE LIST	http://thelist.com/
The World Factbook 1994	http://www.ic.gov/94fact/fb94toc/fb94toc.html
WebCrawler Searching	http://webcrawler.com/
Work Media: Work Software Yellow Pages	http://planetcom.com/workmedia/wsyp.html
World Factbook (CIA)	http://www.ic.gov/94fact/fb94toc/fb94toc.html
Yahoo - Health:Medicine:Cancer	http://www.yahoo.com/Health/Medicine/Cancer/
Yahoo:Health:Medicine	http://akebono.stanford.edu/yahoo/Health/Medicine/

ated by the National Library of Medicine, automatically links abstracts returned in searches with publishers' sites that provide full text versions of their journals. Recently, Harold Varmus, director of the United States National Institutes of Health, described the potential advantages of electronic publishing: open access, rapid and wide dissemination, reduced costs, and flexible publication formats[12] creating healthy debate among the medical establishment.[13] To expand access to research from the developing world, *The Lancet* is proposing an Electronic Research Archive and e-print server for international health articles that may be openly peer reviewed on the Internet by anyone prior to publication in *The Lancet* or just made available as a non–peer reviewed submission that can be updated at the discretion of the author.[14]

The Internet is acting as a potent catalyst for scientific advances by permitting new ties among researchers at any world-wide site. Colleagues can, with permission, be linked directly or by e-mail or fax to any computer on the Internet. Mountains of data, large electronic meetings, bulletin boards, where questions can be answered almost instantaneously, and electronic journal publications have become commonplace. The reader is referred to the bibliography for useful reading sources and to Table 159.3 for Internet WWW sites that offer informatic, medical, and oncologic information,[15,16]

Internet portals are sites that consolidate, organize, and report data, assist with searching for data using search engines, or assist with e-commerce and e-business and are fast achieving the status of an industrial utility. Health-care portals can be divided into three categories. It is helpful to understand them now and to project where they are headed. First are the content-oriented sites, which divide nicely into four sub-groups: patient-centric (like DrKoop.com}, provider-centric (like Healtheon.com), hybrids like webbed (WebMD.com), and companies that have a product to sell like ADAM Software's dissectable database of the human body (adam.com). These companies compare with publishers and information media. Second are the commerce sites, including Internet drugstores and global health-care marketplaces which offer medical products and services. These companies compare with stores. Third are the transaction-based companies like Healtheon that plan to reduce health-care transaction costs by linking key health-care constituencies, such as payers, providers, consumers, diagnostic agencies and suppliers providing patient enrollment, benefits administration, referral claims and tracking, third-party transactions, electronic prescriptions and reporting, and pharmacy benefit management. Listed in Table 159.4 is an index of health-care and general portals that serve as jumping-off points for the Internet.

Table 159.4. Internet Portal Search Engines

Site	Description	URL
Alta-Vista	Fast free-text searching	http://www.altavista.com
AOL NetFind	Searches Web sites, phone numbers, business information, maps and directions, e-mail addresses, newsgroups, personal web sites.	http://www.aol.com/netfind
Dogpile	Searches multiple search engines	http://www.dogpile.com
Euroseek	European source	http://www.euroseek.com
Excite	Concept searching	http://www.excite.com
Google	Free-text searching adding a page's popularity from Sanford	http://www.google.com
Hotbot	Searches multiple search engines simultaneously	http://www.hotbot.com
Infoseek		http://www.infoseek.com
Lycos		http://www.lycos.com
Northern Light	Free-text searches are categorized	http://www.northernlight.com
Profusion	Searches major search engines and filters the results	http://www.profusion.com
Snap		http://www.snap.com
Webcrawler	Metasearch engine	http://www.webcrawler.com
Yahoo	Subject guide and ree-text searching	http://www.yahoo.com

Because of the vastness, reliability, and ephemeral nature of Internet medical information, every oncologist needs to have an understanding of the basic issues related to information available on the Internet. It is important to know the expertise, conflicts of interests, and biases of the developers and sponsors of a site, also, what the site's purpose and intended audience are. Is the original source of content identified, current and valid? Is the user's confidentiality protected, and if not, who has access to information about the users?

Page description languages define the appearance and user interface of the browser screen and are derived from Standard Generalized Markup Language (SGML). This is a meta language used by printers who originally scribbled notes on documents to instruct the typesetters and later codified these shortcuts for universal use. Hypertext Markup Language (HTML) was created as a computer-independent platform for the Internet, using tags to indicate the visual appearance and placement of text and pictures. HTML uses angle-bracketed label tags that surround the text to which they apply and can be nested inside one another to multiple levels. For example, "<BOLD> Name</BOLD>" would produce "Name" on the screen. Although HTML is the most successful electronic-publishing language to date, it can only describe the arrangement of text, images, and push buttons on a page. The deficiencies seen with HTML have all been rectified by the invention of Extensible Markup Language (XML), which combines more efficient processing and flexible linking.[17]

XML permits information to be self-describing, that is, what the information is, not just what it looks like. On the WWW, XML tags might tell the computer that a list of numbers represent the state of your bank account. Thus, the computer will know what the things are, how they are related, and how they should be dealt with. This opens up the possibility of facilitating the searching for and exchanging of scientific data, commercial products, and multi-lingual documents. It tells the computer how text, images, and video are placed on the screen, irrespective of whether the computer is a desktop PC or a hand-held device. Thus, the oncologist could pull up a list of medications that the patient is taking and e-mail the list to a colleague, who could then paste the list into her hospital's own database record for this patient.

Like HTML, XML uses tags that must precede and follow the item described, like parentheses; and like quotation marks, tag pairs can be nested inside one another to multiple levels like a tree structure used to designate hierarchical relationships. To indicate that a patient has a drug allergy to a given drug, the following XML notation would be used: <patient> <name> xxx </name> <drug-allergy>xxx </drug-allergy> </patient>. There are four benefits that XML will offer: (1) a better way to search: Today, a keyword search can return thousands of possibilities. XML tags will filter data and return only the results desired; (2) a better way to distribute and track information: today, it is difficult to republish content across many sites, and more difficult to track who is reading it. XML will make both feasible; (3) a better way to do business: today, one can browse through catalogues online. XML tags will allow data to be customized;and (4) a better way to do business on the road: today, WWW graphics bog down and slow Internet connections. With XML your notebook will download only material tagged as text. XML permits specialized markup languages, such as terms for medical record fields in a database or a chemical markup language that graphically renders a molecular structure when a markup term is used. It also allows a choice of multiple linked sites and automatic updating of all links to a given site when that site changes location. Thus, XML will permit a more powerful and useful WWW that will have great value in simplifying medical information management.

Messaging standards allow applications on the WWW to exchange information in a consistent and secure way. Computers have the capability to send the same information to individuals either in e-mail, fax, pager, or voice-mail format. E-mail can exchange letters and documents from within a word-processing application. Directory services allow a common point of information storage for access by groups of users. Authentication insures that all communications are kept secure and private by using encryption techniques. Digital signatures attach a reliable approval signature to documents. Thus, documents can be securely handled, reviewed, and approved among work groups, facilitating scientific collaboration.

INFORMATION MANAGEMENT

Medical Education Knowledge can be defined as the organization, analysis, and interpretation of data and observations that convey a higher degree of understanding. It is imperative that oncologists have direct access to knowledge for problem solving, and the ability to browse for curiosity's sake and to annotate and personalize knowledge. The major sources of oncologic information are becoming computerized. These include bibliographic information, such as CANCERLIT; factual information, such as PDQ; full text, such as this book; and knowledge bases, such as Oncocin.

Computerized Patient Records The computer-based patient record (CPR) was coined by a committee of the Institute of Medicine of the National Academy of Sciences (NAS) in 1989 to improve the patient record. Oncology with its emphasis on clinical trials and longitudinal data for its patients would seem to be a natural resource for automated records. Within 5 years, all oncologists will be participating in the use of an electronic medical record which will be routinely deployed

Table 159.5. Computer-Based Patient Record Attributes

1. Supports a problem list.
2. Has the ability to measure health status and functional levels.
3. Can document clinical reasoning and rationale.
4. Is a longitudinal CPR and has timely linkages with other patient records.
5. Guarantees confidentiality, privacy, and audit trails.
6. Offers continuous access for authorized users.
7. Supports simultaneously multiple user views into the CPR.
8. Supports timely access to local and remote information resources.
9. Facilitates clinical problem solving.
10. Supports direct data entry by physicians.
11. Supports practitioners in measuring costs and improving quality.
12. Has the flexibility to support existing or evolving needs of clinical specialties.

across all health-care settings. The oncologist will be able to seamlessly create, maintain, edit, display, and manipulate all the data stored in the individual's record. Aggregates of data will be stored in a clinical data repository (CDR), which will serve as the storage facility database for CPRs, facilitating health research and clinical trials.

The Institute of Medicine study listed 12 key attributes of the CPR that has become the gold standard by which all CPRs are judged. These are listed in Table 159.5. In addition to these 12 attributes, more recent underpinnings have been identified as enablers of a robust CPR. These include the need for a minimum standardized clinical data dictionary to define the data elements required in the CPR and the need to support a controlled vocabulary, which defines precisely how a term is used so that there is a uniform understanding of its meaning in clinical medicine. Further provisions will require that ambulatory medical records be created in our offices and be integrated seamlessly with our hospital information systems, such that there is a virtual longitudinal record. Most likely, the set of technologies that make up the Internet will be used to provide the foundation for creating the CPR which will be shared by the patient, provider, and payer, with appropriate encryption, security, confidentiality, and audit trails.

Fortunately, clinical message sending is being standardized around the Health Level Seven (HL7) version 3 standard, and XML. HL7 permits the transfer of information that is vendor-, application- and foreign language–neutral.

As oncologists become more focused on cancer screening and prevention and as data management of clinical trials become automated, disease management will become an accepted part of our practice management, and the CPR will become indispensable. We will see collaborative online efforts to develop clinical pathways (multi-disciplinary plans of care that standardize patient treatment) and bring decision support to the oncologist at the point of care. This will assist us in practicing more evidence-based medicine (bringing current scientific knowledge to the oncologist for point-of-care decisionmaking).

Clinical Trials To increase the number of patients accrued (now less than 3%) and to simplify the data management requirements for cancer clinical trials, two recent innovations deserve comment. The National Cancer Institute's Cancer Informatics Infrastructure (CII) will permit oncologists and their patients to use their computers over the Internet to identify appropriate cancer clinical trials for their specific cancer diagnosis and to facilitate patient registration (*http://cancertrials.nci.nih.gov*).

All phase I, II, and III cancer clinical trials approved by the NCI's Cancer Therapy Evaluation Program (CTEP) are now required to be reported to CTEP using the Clinical Data Update System (CDUS) that is based on coding standards approved by the International Conference on Harmonization Multidisciplinary Group 2. (*www.ich.org*). This group was established to recommend electronic standards for the transfer of regulatory information by evaluating and recommending open and nonproprietary standards that meet the requirements of pharmaceutical companies and regulatory authorities. The recommendations provide solutions for structured and personal messaging (free text), electronic data interchange (EDI), data sets definitions which incorporate structured data formats, security to ensure confidentiality, data integrity, authentication, and nonrepudiation. The data sets include protocol administrative information, patient demographics, treatment information, toxicity information, and response information. Further information about CTEP is available at *http://ctep.info.nih.gov* and for CDUS at *http://ctep.info.nih.gov/ctepinformatics/cdus/default.htm*.

Clinical Practice Guidelines Clinical practice guidelines are systematically developed statements to assist the practitioner with clinical decisions about appropriate health-care for specific clinical indications. With the upsurge in evidence-based medicine and systematic reviews, they have emerged as a health-research tool to analyze and report patterns of care and clinical outcomes. In oncology, the American Society of Clinical Oncology (ASCO) (*www.asco.org*) has concentrated on producing practice guidelines for expensive therapies (hematopoietic colony-stimulating-factors) and surveillance guidelines for breast and colorectal cancer, while the National Comprehensive Cancer Network (NCCN) has developed 28 specific guidelines for oncologic diseases covering 93% of all tumor types and support modalities.[1] The NCCN guidelines include screening, prevention, diagnosis, workup, treatment, and follow-up components. The goal of these techniques is to provide a means to measure an improvement in clinical outcome, in terms of survival, quality of life, treatment toxicity, patient satisfaction, and cost-effectiveness. Clinical practice guidelines will be incorporated into point-of-care oncologic practice by integrating them into oncologic electronic patient records, permitting analysis of patterns of care. In addition, computers will give creators of guidelines instant feedback from users when guidelines are not followed so that guidelines can be constantly assessed.

Information Refinement Computer technology today can increase accessibility to a collection of information. Tomorrow, it will make this information much more useful. Information refining is an electronic, computer-based process that takes undifferentiated volumes of raw information and converts them into electronic form, extracts the content units, and recombines them into a new form that can be distributed in a variety of ways or recombined into new end products. Primitive refining today uses the literal terms of words, keywords, and synonym lists that rely on word indexes and the searching and indexing of keyword bases. It suffers from poor recall performance and poor precision. Future systems will be able to break all information into usable divisions, strip out the impurities, and turn the data into a basic form that can be reprocessed, refined, and transformed.

Unified Medical Language Systsem Medical informatic standards will be required to simplify information management. The Unified Medical Language System (UMLS) project, undertaken by the National Library of Medicine, is designed to facilitate the retrieval and integration of information from many machine-readable information sources, including descriptions of the biomedical literature, clinical records, factual data banks, and medical knowledge bases.[18,19]

Like MESH (Medical Subject Heading), which is a controlled vocabulary of 16,000 terms culled from the medical literature and arranged in a series of tree structures, the proposed UMLS is a highly structured language designed for a clinical database to define medical terminology more specifically. Thus, different software programs that describe signs and symptoms of a disease will use a common term and permit one program to communicate more easily with another. To the extent that an application conforms to the UMLS guidelines, it will be able to communicate with other UMLS systems. An example is the development of the Current Disease Descriptions, which, thus far, have taken 900 diseases out of a possible 4,000 to 5,000 and described them in a "structured text" format, using words and short phrases under labeled contexts.[20]

Clinical Laboratory and Testing Standards A standard for exchange of laboratory data that will standardize the way all medical laboratories represent, store, and transmit data electronically has been defined by the American Society for Testing and Materials (ASTM).[21] Hospitals, the ASTM, and vendors have broadened these standards for

Table 159.6. Boolean Searching Examples

AND: Find all references to A AND B will bring all the records that contain A which also contain B. Records containing only A or only B are not included. It is used to limit a search (reduce the number of records found).
OR: Find all references to A OR B will bring all records that contain A, plus all the records that contain B, plus all the records that contain both A and B. It is used to expand a search (increase the number of records found).
AND NOT: Find all references to A AND NOT B will exclude topics that are not of interest. This means that all records that contain A but which do not contain B should be obtained.
WITHIN xx WORDS: (Proximity Search) Find liver within three words of transplant will bring liver transplant, not transplant of the cornea or marrow.

similar handling of data for billing, admissions, and discharges (ADT), orders, and laboratory reports in hospitals under the HL7 guidelines. In addition, the American College of Radiology (ACR) and the National Electronic Manufacturers Association (NEMA) have formalized the storing of digital radiographic images. This allows hospital, commercial, and physician office laboratories to communicate with each other and facilitate placing information on a patient's chart.

Test Retrieval Retrieval of information from computer databases utilizes either an exact match of terms or phrases (keyword in context or "kwic") or a probabilistic scheme based on the appearance of terms or phrases.[22,23] Only exact matches are economically feasible with today's software. The combination of terms or phrases with the Boolean "AND," "OR," and "NOT" operations, together with adjacency or nearness of terms, serves as the basis for matching against a dictionary all the terms in the database, excluding certain STOP words, such as "the," "and," and "of" (Table 159.6). Keywords, such as MESH terms in MEDLARS, are assigned to citations by abstractors to help facilitate searching, since language is imprecise and exact terms may change with time. The oncologist will find invaluable a recent review by Lowe and Barnett on understanding and using the MESH vocabulary.[24] Full-text databases use proximity searching or the spatial relationships of words and phrases, along with synonyms to facilitate searching.

Electronic databases today use the Boolean algorithm searches. Although they reduce imprecision (the portion of retrieved items that are actually of interest) and increase recall (the portion of truly relevant items that are found), they have only modest memory. Future programs will be able to take a search query written in English, do a spelling check, remove words with low information content, extract root words from the words' different forms and create suffix variants, add synonyms from a thesaurus or MESH dictionary, deliver the articles in the order of importance, and reorder and redo the search when articles of low relevance are indicated by the searcher and the concept words are better defined. Finally, the articles may be presented in the decreasing order of the number of times the context words are repeated in the article.

MAJOR MEDICAL DATABASES OF INTEREST TO THE ONCOLOGIST

The reader is referred to DIRLINE (Directory of Information Resources Online) produced by the National Library of Medicine and available online at *http://www.nlm.nih.gov/pubs/factsheets/dirlinfs* for location and descriptive information about a wide variety of information resources, including organizations, research resources, projects, and databases concerned with health and biomedicine. It is essential to be able to measure cancer-related risk factors, health behaviors, health services, clinical outcomes, and cost from cancer incidence, morbidity, mortality, and survival statistics. The information and computer technologies combining statistical methodologies and national databases evaluating these parameters are readily available for analysis and are available on the WWW. The NCI-coordinated Surveillance, Epidemiology, and End Results Program (SEER) is a population-based registry of cancer registry information from 13 states collecting data on cancer incidence and mortality (*www-seer.ims.nci.nih.gov/*). Population-level data on health behaviors, medical examinations, and health outcomes can be obtained from the National Center for Health Statistics (*www.cdc.gov/nchswww/*). Cost data can be obtained from the Health Care Financing Administration database on medical care and Medicare claims (*www.hcfa.gov/*). The National Cancer Database collected by the American College of Surgeons Committee on Cancer and the American Cancer Society measures patterns of cancer care in the United States (*http://www.facs.org/about_college/acsdept/cancer_dept/programs/ncdb/ncdb.html*). Fraumeni and Rimer have inaugurated a series on the developing fields of cancer surveillance which utilizes these "geospatial databases," a term coined by them, to measure the progress of the National Cancer Program and to provide leads into the etiology and eventual control of cancer.[25]

More detailed summaries of the most popular medical databases are given below.

Popular Databases

Cancernet. CANCERNET (*http://cancernet.nci.nih.gov/*) contains monthly updated information on cancer including PDQ®, fact sheets, publications, and NCI news covering cancer treatment, detection, screening, prevention, rehabilitation, and quality-of-life issues, detailed information about ongoing clinical trials, selected information from the *Journal of the National Cancer Institute*, CANCERLIT® topic searches, links to other NCI and government web sites, and links to other organizations that provide cancer information and support services. In addition, summaries and fact sheets that have been added, changed, or deleted each month can be accessed for the oncologist's review.

Cancerlit. The CANCERLIT database compiled since 1976 contains over 1,400,000 citations from the world's cancer literature beginning from 1963 and with over 60% containing abstracts. About 4,000 medical resources are currently indexed in CANCERLIT, including journals, meeting proceedings, symposia, books, doctoral theses, and research project summaries. Each year an additional 96,000 records are added. Each article is referenced with appropriate MESH terms.

Medline. The MEDLINE database, compiled since 1966, contains over 9,200,000 citations from the world's biomedical literature, with over 60% containing abstracts since 1975 and 80% referring to English language items. About 3,800 of the world's 22,000 medical journals are currently indexed in MEDLINE. Each year 372,000 articles are added. Each article is referenced with carefully chosen key words and phrases from the National Library of Medicine's MESH terms so that one can call up a list of relevant titles and then obtain the most appropriate abstracts. Two popular Internet locations for MEDLINE include Grateful Med *(http://igm.nlm.nih.gov/*) and PubMed (*http://www.ncbi.nlm.nih.gov/PubMed/*).

Aidsline. Produced by the National Library of Medicine since 1980, AIDSLINE contains all medical sources of information on AIDS including citations covering research, clinical aspects, and health policy issues. It is derived from the MEDLINE®, HealthSTAR, and BIOETHICSLINE files. Meeting abstracts from international conferences on AIDS and other AIDS-related meetings, conferences, and symposia are included. There are currently 156,000 records available, and 900 more are added each month (*http://igm.nlm.nih.gov/*).

Biosis. PREVIEWS. Produced by Biosciences Information Services, this database contains six million citations from 1969 to the present, covering some 9,000 scientific journals and books, conferences, proceedings, and monographs.

Embase. This is EXCERPTA MEDICA's online database produced by Elsevier, which screens 4,500 journals published since 1974, making available over four million citations.

HealthSTAR. This database includes patient outcomes and the effectiveness of procedures, programs, products, services, and

Table 159.7. PDQ Selections

1. Cancer Information
    - Treatment by body system/site
    - Treatment by histologic tissue/type
    - Treatment of childhood cancer
    - Supportive care
    - Rare tumor
    - Early detection guidelines
    - Design of clinical trials
    - Cancer screening and prevention
2. Late-Breaking News in Oncology
3. Physicians
4. Organizations
5. Protocols
    - Study objectives
    - Patient eligibility
    - Treatment regimens
    - Dose and schedule modifications
    - Names of principal investigators and institutions
6. Investigational Drug File
7. CANCERLIT Searches

processes and nonclinical health-care administration and planning aspects of health-care delivery. It has 3.1 million records and adds 17,000 per month (*http://igm.nlm.nih.gov/*).

National Cancer Database. Produced by the Commission on Cancer of the American College of Surgeons and the American Cancer Society, this database collects cancer registry data from hospitals and helps oncologists compare trends in the treatment of cancer and its outcomes by individual hospital or by state, regional or national patterns of care. (http://www.facs.org/about_college/acsdept/cancer_dept/programs/ncdb/ncdb.html).

SCISEARCH/Social SCISEARCH. Developed by the Institute for Scientific Information, this resource uses the *Science Citation Index* and the *Social Science Citation Index* to determine the frequency of literature citations for a given article and permits the tracking of all articles that cite common references in order to trace common work.

Physician Data Query. Physician Data Query (PDQ) contains information that has been evaluated by a panel of 70 experts and is updated on a quarterly basis. It gives detailed summaries for the physician or layperson of the 80 major tumor types, including prognosis, staging, cellular classification, and state-of-the-art treatment options. Tumors may be searched by body system/site or histologic tissue/type. It lists over 1,500 currently active protocols that use mostly standardized formats and 6,000 closed-treatment protocols, with detailed information on the study objectives, patient entry criteria, treatment regimens, and demographic information for cooperative groups and physicians involved with the protocols. In addition, PDQ makes available demographic information for approximately 13,000 cancer specialists and over 1,500 organizations affiliated with societies that are related to cancer. It is readily available to physicians and their patients through telecommunication facilities or telephone (1-800-4-CANCER). The reader is referred to two excellent reviews for further information.[26,27] Table 159.7 lists the major categories of PDQ.

PDQ is primarily used by oncologists to ensure the use of the most current treatment methods, to obtain help to make a clinical decision, and to find information about clinical trials. Less frequently, it is used to prepare for conferences or to confirm information from other sources. Other uses include referring a patient to an oncologic investigator in a specific location, finding an investigator who has access to a particular clinical trial, and seeking consultation from another physician.

SEER. The NCI's Surveillance, Epidemiology and End Results (SEER) tumor registry of cancer cases began in 1973 and collects cancer demographic information from nine geographic areas in the United States. Researchers plan to combine this database with the Health Care Financing Administration's (HCFA) Medicare claims files for the same patients. This will help determine regional cancer treatment and staging procedure costs and the cost-effectiveness of screening and staging procedures and to compare hospital and outpatient costs of cancer care to assist in outcome analyses and practice guidelines research.

CCRIS. The Chemical Carcinogenesis Research Information System (CCRIS) contains carcinogenicity, tumor promotion, tumor inhibition, and mutagenicity test results derived from the scanning of primary journals, current awareness tools, NCI technical reports, review articles, and International Agency for Research on Cancer monographs published since 1976 (*http://toxnet/nlm.nih.gov*).

CLINICAL AND RESEARCH USES OF COMPUTERS IN MEDICAL ONCOLOGY

Computer-generated Media Computers integrate text, graphics, animation, audio, and video communication technologies into a potent method to deliver information. Oncologists now have a choice of media—print, slides, overhead, or live delivery from a computer screen. An idea can be drawn on paper and scanned into the computer, or drawn directly on the computer screen using up to 32 million colors, and even animated and given sound. Presentations using electronic links can have the input of many individuals and can be instantly altered for different audiences. Using a presentation program, like PowerPoint from Microsoft, for controlling external devices, such as a computer projector, one can integrate multimedia for live presentations.

Computers are revolutionizing medical arts departments in medical schools. Oncologists can easily create graphics of their data and ideas and transmit them to an audio or video device. Changes can be reflected instantaneously. Laser printers can provide almost typeset-quality reproductions. Scanners replace paper images, such as photographs, illustrations, text, files, forms, and charts, with digitized images of the original, which, when converted into a computer document, can be customized for presentation. These images can be stored on magnetic media and transmitted over data communication networks. Thus, users can scan, store, retrieve, enhance, sort, and route these documents.

Document Imaging and Image Processing Because medical imaging plays such a major role in the oncologist's need to measure and define cancer, a method to acquire, display, analyze, store, and print digitized images becomes important. Radiographs, nuclear medicine scans, CT and magnetic resonance imaging (MRI) scans, and motion and still video films can all be viewed on the high-resolution screens of personal computers so as to display images with the same clarity, resolution, and gray scale as the imaging systems' own consoles. Modern compression techniques can store up to 10,000 images on an optical disk for instantaneous retrieval for patient care, teaching, and documentation.

Document imaging is the replacement of paper images and files with digitized images of the original, using a scanner. These images are stored on magnetic media and can be transmitted over data communication networks. However, a price is paid for this technique, because of the large requirement for memory. An 8.5×11 page requires 1,000K if stored as a 300-dot-per-inch picture (8.5×11 inches $\times 300^2$ dots per inch = 8,415,000 bits or 8,415,000/8 = 1.05 megabytes) compared with 1.5K, if stored as a text file. However, compression techniques can store a typewritten page at 150-bit-per-inch resolution in only 20K of memory. The digital transmission of colored television full-motion audio and video requires 45 million bits per second, and even a still TV image requires 24 million bits of memory.

CT, MRI, and ultrasonography have remained the province of the radiologist. But now, from their desktop computers, oncologists will be able to obtain these images; enhance, store, index, and transfer them to any computer worldwide; and manipulate them to make diagnosis and treatment planning easier. The computer is able to compare changes on different studies and give quantitative and qualitative information more accurately than can the human eye. Radiographs are now digitized for similar analyses.

Three-dimensional imaging has revolutionized radiation treatment planning. Two-dimensional digitized medical CT or MRI images can be combined with three-dimensional graphic programs to target tumor killing and minimize damage to normal tissue. By defining tumor location and anatomy, the computer can then design the proper radiation-beam approaches and calculate the radiation dose distribution. Similar techniques will make surgical treatment more exacting. NIH Image is a public-domain (free) image acquisition, measurement, and processing program (developed by Wayne Rasband of NIH) that reads all common graphic files and supports many standard image-processing functions and has been widely adopted by scientists for medical imaging applications. It is available on the Internet by anonymous file transfer protocol (FTP) from *zippy.nimh.nih.gov*.

Presentation Devices Output devices permit us to use presentation software to make 35-mm slides for display via conventional slide projectors. With inexpensive projection pads, they also permit enlarging the computer screen by using an overhead projector. Animated and audio presentations can be produced for educational and instructional purposes; they can be transmitted to a television or video recorder. Cancer patients who have disabilities can use voice recognition software to control the computer. Textual material can be converted to speech by specialized software to assist the visually impaired.

Scanners Black-and-white or color scanners can convert photographs, line-art drawings, and gray-scale or halftone images into digital form for manipulation, visualization, and storage in the computer. These can then be used directly in a word-processing or page makeup program or placed in a presentation program for slides. Scanners can even take radiographic films and scan them into the computer.

Optical Character Recognition Optical character recognition (OCR) software can, with almost 100% accuracy, take typewritten or typeset text created with any font or font size, in any language; distinguish text from graphics and multiple columns of text in a table; and convert it into a digital form that can be manipulated by the computer. A scanner can easily put typewritten reports into an electronic medical record. A single typewritten page can be scanned in less than 10 seconds. Thus, large amounts of printed or faxed material can be easily and inexpensively stored, indexed, tracked, and made easily accessible.

Presentation Software Organization, outlining, graphic skills, and visual presentation are required to convey our thoughts to our colleagues. Presentation software can assist in creating, organizing, and producing presentations with the capability of easily making modifications for different audiences. Outliners can organize text into an outline form and transform it into bullet charts or organizational charts for on-screen or overhead slide presentations. Presentation programs control the font, size, style, and color of individual characters. One can scan into the computer diagrams, tables, graphs, or pictures; add text; modify the drawing; sort the slides; produce speaker notes; and then send the information to a desktop film recorder for slides, to a color ink-jet printer for handouts or overhead slides, or to service bureaus for professionally produced slides. One can even directly present what is seen on the computer monitor on a large screen, using a projection pad and an overhead projector.

Statistics There is a plethora of statistical programs that combine the major procedures used in most analyses with the ability to display results instantly as charts or plots. Most programs permit the data to be read in from other programs and to be stored in tables, rectangular files, or matrices. After analysis, the results can be exported to other programs. Data can be sorted, selected, and rotated on the basis of data values. Most programs include frequency distribution; data transformation; nonparametric statistics; cross-tabulation statistics; analysis of variance, correlation, and regression; and multivariate analysis.

Data Acquisition and Instrument Control Data acquisition requires taking measured experimental data in wavelengths or analog form and converting the data to digital form by an analog-to-digital-converter. With specialized computer software, it is possible to have the computer read this information, display it, analyze it, and make appropriate decisions automatically. Sound, heat, light, pressure, and electrical signals can be recorded by the computer, in effect, transforming it into a voltmeter, frequency counter, oscilloscope, spectrum analyzer, or x-y recorder. The most popular instrument control program is Lab View (National Instruments Corp., Austin, TX).

Molecular Biology and Gene Sequencing Using homology matrix analysis, the computer can compare sequences of numbers or letters. It, thus, can compare DNA sequences and perform DNA sequencing. Such an innovative program is Gene Construction Kit for Macintosh by Bob Gross of Dartmouth (Textco Inc., West Lebanon, NH). Users can represent DNA sequences graphically or as text. Recombinant DNA molecules can be constructed on the screen. The program has a database of restriction enzymes that can cut DNA at specific base pairs. Thus, one can cut and splice DNA sequences and then analyze the results in the model. It is a means to bring the world of desktop publishing into the molecular biologist's world.

ONCOLOGY OFFICE AND HOSPITAL INFORMATION SYSTEMS

Clinical Data Systems A clinical data system is a computer-based system that captures, stores, or communicates clinical data to enhance clinical decision making. Clinical information combines specific patient data (history or laboratory data, for example) and knowledge (facts about a cancer or how to interpret a laboratory result) that applies to many patients. Computers are essential for both types of data, with a computer-based patient record holding patient data and clinical databases and electronic medical textbooks containing clinical knowledge. Wyatt breaks down clinical data into five categories: identifiers (demographic data, identity numbers); patient findings (history and observations, including symptoms, clinical signs, and laboratory results); hypotheses (assessment and plan, such as diagnosis, stage, problem list, and possible tests or therapies); actions (therapy, referral, tests ordered, information given); and modifiers (who recorded data, when, certainty).[28] He identifies free text, diagrams, text, numerical report, graphs, charts, photographs, and tracings as the different forms of clinical data that can be assimilated by the computer. The appearance of clinical data takes on added importance because of the ease of viewing and arranging data by the computer. Powsner and Tufte proposed using multiple small graphs printed out on a laser printer to summarize a mass of clinical and laboratory data from a patient's clinical chart.[29] Numerical data can then be plotted to reveal the course of an illness or the patient'sresponse to treatment.

Wirtschafter and colleagues implemented a computerized consultant extender for a breast cancer clinical protocol that used algorithms to produce a 95% compliance rate among 75 local physicians who used the system, compared with a 64% rate for cancer center physicians who did not use the automated system.[30] ONCOCIN is an expert system, in use at Stanford University for assisting the oncologist to manage patients on chemotherapy protocols.[31,32] It is both a medical record and an advice system that records the course of the patient on a protocol and can assume an active role in adapting the protocol to the individual patient. It can determine drug doses on the basis of time schedule, toxicity, and blood counts. When oncology patient care stipulated by the computer was compared with care given by oncology fellows, similar levels of competency were seen.

Numerous oncology medical records have been implemented at cancer centers to facilitate all aspects of patient care, to interact with hospital information systems and with data management offices, and to maintain patient and protocol charts. The CLIN system implemented at the Memorial Sloan-Kettering Cancer Center by Serber was one of the earliest systems that implemented an oncology protocol record with complete data management and statistical tools.[33] The OCIS (Oncology Clinical Information System) of the Johns Hopkins Hospital is a time-oriented tabular and graphic database with a meaningful grouping of related information that links patient data and procedural support.[34] It interconnects mainframe, minicomputers, and microcomputers effectively to automate all aspects of cancer patient management.

Many oncology practices are becoming computerized. Numerous programs are available for obtaining medical and scientific information, and for office administration, accounts receivable and payable, appointment scheduling, electronic banking and claim forms, and financial and practice management. The reader is referred to the *15th Annual Medical Hardware and Software Buyer's Guide,* which lists over 1,000 medical software products, the *1999 Resource Guide from Healthcare Informatics; Software for Internists;* and the monthly newsletter *Medical Software Reviews.*[35–38]

All oncologists must become computer literate. A small investment in learning time and capital ($1,200 to $3,000) will pay off handsomely. Oncologists will have easier access to patient data, less loss of records, greater medical knowledge, and more efficient updating of information. They would have immediate access to electronic textbooks and journals and easier communication from one's own office, using electronic mailboxes and modems, with other physicians' offices or the hospital. The computer has become the oncologist's "black bag."

REFERENCES

1. Winn RJ. National Comprehensive Cancer Network Oncology Practice Guidelines. Supp Oncol 1996;1–317.
2. Gotay CC. Trial-related quality of life: using quality-of-life assessment to distinguish among cancer therapies. Monogr Natl Cancer Inst 1996;20:1–6.
3. Weeks J. Taking quality of life into account in health economic analyses. Monogr Natl Cancer Inst 1996;20:23–27.
4. Greenes RA, Shortliffe EH. Medical informatics: an emerging academic discipline and institutional priority. JAMA 1990;263:1114–1120.
5. Shortliffe EH, Perreault LE. Medical informatics: computer applications in health care. Reading, MA: Addison-Wesley; 1990.
6. Kassirer JP. The next transformation in the delivery of health care [editorial]. N Engl J Med 1995;332:52–54.
7. Gilder G. The coming software shift. Forbes ASAP 1995; 147–162.
8. De Solla Price D. The development and structure of the biomedical literature. In Coping with the biomedical literature. Warren K, editor. New York: Praeger; 1981.
9. Toong HD, Gupta A. Personal computers. Sci Am 1984;254:87–106.

10. Glowniak JV, Bushway MK. Computer networks as a medical resource. JAMA 1994;271:1934–1939.
11. Lawrence S, Giles CL. Search engine. Nature 1999;400:107–109.
12. Varmus H. E-Biomed: a proposal for electronic publications in the biomedical sciences. *www.nih.gov/welcome/director/ebiomed/ebi.html.*
13. Relman AS. The NIH "E-BIOMED" proposal; a potential threat to the evaluation and orderly dissemination of new clinical studies. N Engl J Med 1999;340:1793–1794.
14. McConnell J, Horton R. Electronic research archive in international health and e-print server. Lancet 1999;354:2–3.
15. Hahn H, Stout R. The internet complete reference. Berkeley, CA: Osborne-McGraw-Hill; 1994.
16. Marine A, Kirkpatrick S, Neou V, Ward C. Internet: getting started. Englewood Cliffs, NJ:Prentice-Hall; 1993.
17. Bosak J, Bray T. XML and the second-generation WEB. Sci Am 1999;269:89–93.
18. Humphreys BL, Lindberg DAB. Building the unified medical language system. In: Proceedings, 13th Annual Symposium on Computer Applications in Medical Care. Kingsland LC III, editor. Washington: IEEE Computer Society Press; 1989. p. 475–480.
19. Tuttle MS, Blois MS, Erlbaum MS, et al. Toward a bio-medical thesaurus: building the foundation of the UMLS. In: Proceedings, 12th Annual Symposium on Computer Applications in Medical Care. Greene RA, editor. Washington: IEEE Computer Society Press, 1988. p. 191–195.
20. Rennels G. A computational model of reasoning from the clinical literature. New York:Springer-Verlag; 1987.
21. McDonald CJ, Hammond WE. Standard formats for electronic transfer of clinical data. Ann Int Med 1989;110:333–335.
22. Belkin NJ, Croft WB. Retrieval techniques. In: Annual review of information science and technology. Williams ME, editor. Amsterdam: Elsevier Science; 1987. p. 22.
23. Fenichel CH. The process of searching online bibliographic data bases: a review of research. Libr Res 1980;2:107.
24. Lowe HJ, Barnett O. Understanding and using the medical subject headings (MESH) vocabulary to perform searches. JAMA 1994;272:1103–1108.
25. Fraumeni JF, Rimer BK. Cancer Surveillance Series: Inauguration. J Natl Cancer Inst 1999;1004.
26. Hubbard SM, Henney JE, DeVita VT Jr. A computer data base for information on cancer treatment. N Engl J Med 1987;316:315–318.
27. Hubbard SM. The physician data query (PDQ) cancer information system. Bull Cancer 1987;74:205.
28. Wyatt JC. Clinical data system, Part 1: Data and medical records; Part 2: Components and techniques; Part 3: Development and evaluation. Lancet 1994;344:1543–1547, 1609–1614, 1682–1688.
29. Powsner SM, Tufte ER. Graphical summary of patient status. Lancet 1994;344: 386–389.
30. Wirtschafter DD, Gams R, Ferguson C, et al. Clinical protocol information system. In: Proceedings, Fourth Annual Symposium on Computer Applications in Medical Care. O'Neill JT, editor. Washington: IEEE Computer Society Press; 1980. p. 745.
31. Rennels GD, Shortliffe EH. Advanced computing for medicine. Sci Am 1987;257: 154–161.
32. Shortliffe EH. Computer programs to support clinical decision making. JAMA 1987;258:61–66.
33. Serber MJ, Mackey R, Young CW. The Sloan-Kettering Information System for clinical oncology. In: Proceedings, Fourth Annual Symposium on Computer Applications in Medical Care. O'Neill JT, editor. Washington: IEEE Computer Society Press; 1980. p. 728.
34. Enterline JP, Lenhard RE Jr, Blum BI. A clinical information system for oncology. New York: Springer-Verlag; 1989.
35. The 15th annual medical hardware and software buyer's guide. MD Comput 1998;11:362–400.
36. 1999 Resource Guide. Healthcare Inform 1998; 17–153.
37. Frisch S. Software for internists. Philadelphia: American College of Physicians; 1994.
38. Frisch S. Medical software reviews. Brooklyn,NY:Healthcare Computing Publications;