Basic concepts in medical informatics

J C Wyatt, J L Y Liu

This glossary defines terms used in the comparatively young science of medical informatics. It is hoped that it will be of interest to both novices and professionals in the field.

Medical informatics is the study and application of methods to improve the management of patient data, clinical knowledge, population data, and other information relevant to patient care and community health. It is a young science, which emerged in the decades after the invention of the digital computer in the 1940s. Mechanical computing in medicine had a much earlier origin, in the 19th century, with Herman Hollerith’s “punched-card data-processing system” originally used for the US census and subsequently developed to support surveys in public health and epidemiology. This example reflects the multidisciplinary nature of medical informatics, which interacts with various fields, including the clinical sciences, the public health sciences (for example, epidemiology and health services research) as well as cognitive, computing, and information sciences.

Given the diversity of backgrounds of medical informatics workers, newcomers can easily be confused by the jargon used in the field. An introduction to basic concepts would therefore be useful to those interested in learning more about medical informatics. In recent years, various branches of the discipline have appeared, including public health informatics, consumer health informatics, and clinical informatics. In a debate on whether medical informatics and these branches are distinct disciplines, Shortliffe and Ozbolt argued that “informatics is built on a reusable and widely applicable set of methods that are common to all health science disciplines, and that ‘medical informatics’ continues to be a useful name for a composite core discipline that should be studied by all students, regardless of their health profession orientation.”

Our definitions of the various branches of medical informatics below reflect this.

One or more of the following criteria were used to select terms included in this list of concepts:

- There is general agreement on what the term means and how it is used.
- There is value in attempting to define the term, even though there may be debate about it.

We trust this list of terms will be informative for novices and encourage discussion among medical informatics professionals, particularly on terms for which there is no widespread agreement. For readers who are interested in pursuing the subject further, additional resources are cited at the end of the list.

**ALGORITHM**
A process for carrying out a complex task broken down into simple decision and action steps. Often assists the requirements analysis process carried out before programming.

**BIOINFORMATICS**
The use of medical informatics methods to facilitate research in molecular biology.

**CHECKLIST**
A type of clinical decision tool: a form listing one or more items of patient data to be collected before, during or after an encounter; can be paper or computer based.

**CLINICAL CODING SYSTEM (CLINICAL THESAURUS, CONTROLLED VOCABULARY)**
A limited list of preferred terms from which the user can draw one or more to express a concept such as patient data, a disease or drug name, etc. An alphanumeric code corresponding to the term is then stored by the computer. Synonyms or close matches to each preferred term are usually available and map onto the same internal codes. This approach makes it easier for a computer to analyse data than the use of free text words or phrases. Examples of clinical coding systems include SNOMED-CT (divergent codes used to capture patient data), MeSH (terms used to index biomedical literature) and ICD-10 (convergent disease codes for international comparisons, with specific rules to guide coders). Clinical coding systems play a key part in epidemiological studies and health service research, from the use of MeSH terms to conduct literature searches for systematic reviews to studies that use ICD codes to classify and compare diseases. To prevent information loss, it is vital that the terms and
codes are never changed or dropped, only added to. Obsolete terms can be marked as such to deter inappropriate use. Continuing maintenance is needed to incorporate new terms and codes for new concepts and synonyms as they arise.

**CLINICAL DATA SYSTEM**
Any information system concerned with the capture, processing, or communication of patient data.

**CLINICAL DECISION TOOL**
Any mechanical, paper, or electronic aid that collects or processes data from an individual patient to generate output that aids clinical decisions during the doctor-patient encounter. Examples include decision support systems, paper or computer reminders and checklists, which are potentially useful tools in public health informatics, as well as other branches of medical informatics.

**CLINICAL INFORMATION**
Organised patient data or medical knowledge used to make clinical decisions (adapted from Shortliffe et al); may also include directory information. Many activities in public health and epidemiology (for example, surveillance systems, cohort studies to assess the effects of a risk factor of disease, and clinical trials to estimate efficacies of new treatments) entail the organisation of such data (for example, case report forms for individual patients) into useable information (for example, incidence of notifiable cases of disease from surveillance programmes and summary evidence from cohort studies or clinical trials, expressed as odds ratios for certain harmful and beneficial outcomes). See also: information.

**CLINICAL INFORMATICS**
The use of medical informatics methods to aid management of patients using an interdisciplinary approach, including the clinical and information sciences.

**COMMUNICATION**
The exchange of information between agents (human or automated) face to face or using paper or electronic media. Requires the use of a shared language and understanding or common ground.

**COMPUTER VISION (IMAGE INTERPRETATION)**
The use of computer techniques to assist in the interpretation of images, such as mammograms.

**CONFIDENTIALITY (PROTECTING PRIVACY)**
The policies restricting access to a person’s data to those whom the patient agrees need access to them, except rarely in emergency and for the public good (for example, to contain epidemics, allow important research to be undertaken, or solve serious crime). In addition, other regulatory and institutional approval may be needed (for example, the need to seek consent from medical ethics committees or relevant national authorities). In recent years, leading public health researchers have warned that legislation enacted to protect patients’ medical data in the UK, Europe, and US could potentially hamper observational research and medical record linkage studies.

**CONSUMER HEALTH INFORMATICS**
The use of medical informatics methods to facilitate the study and development of paper or electronic systems that support public access to and use of health and lifestyle information. For additional discussion on the scope of consumer health informatics, see Eysenbach. See also eHealth.

**DATA QUALITY**
The degree to which data items are accurate, complete, relevant, timely, sufficiently detailed, appropriately represented (for example, consistently coded using a clinical coding system), and retain sufficient contextual information to support decision making.

**DATABASE**
A collection of data in machine readable format organised so that it can be retrieved or processed automatically by computer. A flat file database is organised like a card file, with many records (cards) each including one or more fields (data items). A relational database is organised as one or more related tables, each containing columns and rows. Data are organised in a database according to a schema or data model; some items are often coded using a clinical coding system.

**DECISION SUPPORT SYSTEM (COMPUTER DECISION AID)**
A type of clinical decision tool: a computer system that uses two or more items of patient data to generate case specific or encounter specific advice. Examples include computer risk assessors to estimate cardiovascular disease risk and the Leeds Acute Abdominal Pain system, which aided the diagnosis of conditions causing such pain. Evidence adaptive decision support systems are a type of decision aid with a knowledge base that is constructed from and continually adapts to new research based and practice based evidence.

**DECISION TREE**
A way to model a complex decision process as a tree with branches representing all possible intermediate states or final outcomes of an event. The probabilities of each intermediate state or final outcome and the perceived utilities of each are combined to attach expected utilities to each outcome. The science of drawing decision trees and assessing utilities is called decision analysis.

**DEMONSTRATION STUDY**
Study that establishes a relation—which may be associational or causal—between a set of measured variables. In epidemiology, cohort studies, randomised trials, and blind comparisons of a test with a gold standard are typical demonstration studies. See also measurement studies.

**DIRECTORY INFORMATION**
Information specific to an organisation or service that is useful in managing public health services, health care services, or patients. Examples include a phone directory, a lab handbook listing available tests and specimens to use, and a list of drugs in the local formulary.

**eHealth**
The use of internet technology by the public, health workers, and others to access health and lifestyle information, services and support; it encompasses telemedicine, telecare, etc. For discussion on the scope and security issues of eHealth, see a recent report by the National HealthKey Collaborative.

**ELECTRONIC HEALTH RECORD (EHR)**
In the UK, the lifelong summary of a person’s health episodes, assembled from summaries of individual electronic patient records and other relevant data.

**ELECTRONIC PATIENT RECORD (EPR)**
A computer based clinical data system designed to replace paper patient records.
EXPLICIT KNOWLEDGE
Knowledge that can be communicated on paper or electronically, without person to person contact. Public health workers and physicians cannot use explicit knowledge if they cannot access it. There is thus a need to identify, capture, index, and make available explicit knowledge for professionals, a process called codification. Much of the work done by the Cochrane Collaboration entails codification of explicit knowledge. See also: tacit knowledge.

GEOGRAPHICAL INFORMATION SYSTEM (GIS)
Computer software that captures, stores, processes, and displays location as well as other data. The display may preserve distance ratios between data objects (for example, true scale maps) or link similar objects, ignoring distance (for example, topological maps such as that distributed to the public for the London Underground). GIS software is used in many ecological studies of disease. A famous example is Peto’s study of diet, mortality, and lifestyle in rural China. Disease mapping studies have also been conducted to assess childhood leukaemia in areas with different radon levels, the clustering of respiratory cancer cases in areas with a steel foundry, and socioeconomic gradients in infant mortality. GIS are also used for public health planning and surveillance purposes at local or national health departments. Care should be taken by policy makers in interpreting maps produced by GIS software, particularly in regard to the ecological fallacy.

INFORMATION
Organised data or knowledge used by human and computer agents to reduce uncertainty, take decisions, and guide actions (adapted from Shortliffe et al and Wyatt). See also: clinical information, patient data, medical knowledge.

INFORMATION DESIGN
The science and practice of designing forms, reports, computer screens, etc, so that the information they contain can be found rapidly and interpreted without error (adapted from Sless). Information design is based on psychological and graphical design theories and empirical studies of human perception and decision making using alternative formats for information.

KNOWLEDGE BASE
A store of knowledge represented explicitly so that a computer can search and reason with it automatically; often uses a clinical coding system to label the concepts. See also decision support system.

KNOWLEDGE BASED SYSTEM (EXPERT SYSTEM)
A computer decision support system with an explicit knowledge base and separate reasoner program that uses this to give advice or interpret data, often patient data.

KNOWLEDGE MANAGEMENT
The identification, mobilisation and use of knowledge to improve decisions and actions. In public health and medicine, much of this work involves the management of medical knowledge (from epidemiological studies, randomised controlled trials, and systematic reviews) so that it is actually used by the physician. This entails practice innovation or narrowing the gap between what we know and what we do. The NHS is developing a program of knowledge codification to inform routine problem solving, for example, through the National Electronic Library of Health, guidelines from the National Institute of Clinical Excellence (NICE), and care pathways and triage algorithms used in the NHS Direct Clinical Advice System.

MEASUREMENT STUDY
Study of the reliability, validity, or ease of use of a measurement instrument or method in a defined population. See also demonstration study.

MEDICAL INFORMATICS (HEALTH INFORMATICS)
The study and application of methods to improve the management of patient data, medical knowledge, population data and other information relevant to patient care and community health. Unlike some other definitions of medical informatics (or example, Greenes and Shortliffe), this definition puts the emphasis on information management rather than technology. Branches of medical informatics include bioinformatics, clinical informatics, consumer health informatics and public health informatics.

MEDICAL KNOWLEDGE (CLINICAL KNOWLEDGE)
Information about diseases, therapies, interpretation of lab tests, etc, which is potentially applicable to decisions about multiple patients and public health policies, unlike patient data. This information should where possible be based on sound evidence from clinical and epidemiological studies, using valid and reliable methods. See also: explicit knowledge, tacit knowledge, knowledge management.

MINIMUM DATASET
A list of the names, definitions and sources of data items needed to support a specific purpose, such as surveillance of the health of a community, investigation of a research hypothesis or monitoring the quality of care in a registry.

OBJECTIVISTIC EVALUATION
An evaluation approach that uses experimental designs and statistical analyses of quantitative data. Such an approach is never completely objective. See also subjectivistic study.

ONTOLOGY
A description of the concepts and relations in a domain, such as drug prescribing. Sample concepts here would be “patient”, “prescriber”, and “drug”; relevant relations might include “prescribes to”, “requests prescription from”, and “causes side effects to”. A taxonomy or hierarchy is a simple kind of ontology in which concepts are arranged according to only one relation: “is a kind of”. Note that ontology as used here has a different meaning from its use in the philosophy of science, an area of interest to theoretical epidemiologists.

PATIENT DATA
Information about an individual patient and potentially relevant to decisions about her current or future health or illness. Patient data should be collected using methods that minimise systematic and random error. See also: medical knowledge, data quality.

PRACTICE INNOVATION (RESEARCH IMPLEMENTATION, BEHAVIOUR CHANGE)
Applying usually explicit knowledge to innovate in public health and clinical practice by identifying barriers to change and applying appropriate practice innovation methods to overcome these (adapted from Wyatt). See also: knowledge management.

PRACTICE INNOVATION METHODS (RESEARCH IMPLEMENTATION, BEHAVIOUR CHANGE METHODS)
A range of methods that can potentially overcome barriers to change in the practice innovation process, such as clinical audit, outreach visits or clinical decision tools.
PUBLIC HEALTH INFORMATICS
The use of medical informatics methods to promote public health practice, research, and learning, using an interdisciplinary approach, including the public health sciences, for example, epidemiology and health services research, and the information sciences, for example, computing science and technology (adapted from Yasnoff et al18). In a recent paper outlining an agenda for developing this branch of informatics, Yasnoff et al10 argued for the need to construct, implement, and integrate public health surveillance systems at national and local levels, to enable rapid identification and response to disease hotspots (and more typically, bioterrorism). As Yasnoff rightly points out, methods of assessing costs and benefits of such systems are needed. Public health informatics can also contribute in other areas, for example, reminders have played an important part in prevention programmes such as smoking cessation advice to smokers16 and the use of preventive care for patients.10

REGISTRY
A database and associated applications that collects a minimum dataset on a specified group of patients (often those with a certain disease or who have undergone a specific procedure), health professionals, organisations, or even clinical trials. Registries can be used to explore and improve the quality of care or to support research, for example to monitor long term outcomes or rare complications of procedures. Key issues in registries are maintaining confidentiality, coverage of the target population, and data quality.

REMEMBER
A type of clinical decision tool that reminds a doctor about some item of patient data or clinical knowledge relevant to an individual patient that they would be expected to know. Can be paper based or computer based; includes checklists, sticky labels on front of notes, an extract from a guideline placed inside notes, or computer based alerts. There has been much interest in reminders as a practice innovation method recently because of the poor uptake of practice guidelines, even those based on good quality evidence. An example is the treatment of dyslipidaemia in primary care, where there is a big gap between recommendations and actual practice.15

REQUIREMENTS ANALYSIS
The process of understanding and capturing user needs, skills, and wishes before developing an information system (adapted from Somerville29). See also software engineering.

SECURITY
The technical methods by which confidentiality is achieved.16

SOFTWARE ENGINEERING
The process of system development, documentation, implementation, and upgrading (adapted from Somerville29). In the classic or “waterfall” model of software engineering, requirements analysis leads to a document that serves as the basis for a system specification and database schema, from which programmers work to develop the software. However, increasingly, users and software designers work together from the start to develop and refine a prototype system. This helps to engage the users, educate the software development team, brings the requirements documents alive, and allows users to explore how their requirements might change as a result of interaction with the new software.

SUBJECTIVISTIC EVALUATION
An evaluation approach that relies primarily on qualitative data derived from observation, interview, and analysis of documents and other artefacts.15 The focus of such studies is on description and explanation; they tend to evolve rather than be prescribed in advance. As we can never truly understand another person’s feelings, such studies always approximate the subjective. See also objectivistic study.

TACIT KNOWLEDGE (INTUITION)
Knowledge that requires person to person contact to transfer and cannot be communicated on paper or electronically.19,27 Over time, some tacit knowledge can be analysed, decomposed, and made explicit. See also: explicit knowledge.

TELECARE
A kind of telemedicine with the patient located in the community (for example, their own home); see also eHealth.

TELEMEDICINE
The use of any electronic medium to mediate or augment clinical consultations. Telemedicine can be simultaneous (for example, telephone, videoconference) or store and forward (for example, an email with an attached image). See also eHealth.

Additional resources
Readers who are interested in general coverage of the field of medical informatics are encouraged to refer to standard texts.11–13 Those who are interested in alternative or complementary definitions of the above terms can look up various sources.6 10 16

ACKNOWLEDGEMENT
We thank Aimen Abu Hanna (Department of Medical Informatics, University of Amsterdam) and the JECH anonymous referees, who all provided useful comments on drafts.

AUTHORS’ AFFILIATIONS
J C Wyatt, Department of Medical Informatics, Academic Medical Centre, University of Amsterdam, Netherlands
J LY Liu, Centre for Statistics in Medicine, Institute of Health Sciences, Oxford University, UK
Funding: the NHS R&D Health Technology Assessment Board funded part of this work.

Conflicts of interest: none.

REFERENCES

www.jech.com