

What Are Clinical Data?

If data are central to all health care, it is because they are crucial to the process of decision making

clinical datum: any single observation of a patient—e.g., a temperature reading, a red blood cell count, a past history of rubella, or a blood pressure reading.

The idea of a data model for computer-stored medical data accordingly becomes an important issue in the design of medical data systems

single datum generally can be viewed as defined by five elements:

1. The patient in question
2. The parameter being observed (e.g., liver size, urine sugar value, history of rheumatic fever, heart size on chest X-ray film)
3. The value of the parameter in question (e.g., weight is 70 kg, temperature is 98.6 °F, profession is steel worker)
4. The time of the observation (e.g., 2:30 A.M. on 14FEB2013 1)
5. The method by which the observation was made

A related issue is the uncertainty in the values of data:

- An adult patient reports a childhood illness
- A physician listens to the heart of an asthmatic child and thinks that she hears a heart murmur—but is not certain because of the patient's loud wheezing
- A radiologist looking at a shadow on a chest X-ray film is not sure whether it represents overlapping blood vessels or a lung tumor.

The idea of trade-offs in data collection thus becomes extremely important in guiding health care decision making.

2.1.1 What Are the Types of Clinical Data?

broad range of data types in the practice of medicine and the allied health sciences. They range from narrative, textual data to numerical measurements, genetic information, recorded signals, drawings, and even photographs or other images

Note that there are significant problems associated with the use of such abbreviations. Many are not standard and can have different meanings depending on the context in which they are used. For example, “MI” can mean “mitral insufficiency” or “myocardial infarction”. Many hospitals try to establish a set of “acceptable” abbreviations with meanings, but the enforcement of such standardization is often unsuccessful.

idea of data is inextricably bound to the idea of data recording

The notations may be highly structured records with brief text or numerical information, hand-drawn sketches, machine generated tracings of analog signals, or photographic images (of the patient or of radiologic or other studies). This range of data- recording conventions presents significant challenges to the person implementing electronic health record systems.

2.1.2 Who Collects the Data?

- health care team evoke images of coworkers treating ill patients
- nurses play a central role in making observations, because nurses typically spend more time with patients than physicians do
- Physical or respiratory therapists Laboratory personnel perform tests on biological samples
- Radiology technicians
- Pharmacists may interview patients about their medications or about drug allergies
- technological devices that generate data—laboratory instruments, imaging machines, monitoring equipment in intensive care units, and measurement devices that take a single reading (such as thermometers, ECG machines, sphygmomanometers, and spirometers for testing lung function)

2.2 Uses of Health Data:

2.2.1 Create the Basis for the Historical Record

- What is the patient's history (development of a current illness; other diseases that coexist or have resolved; pertinent family, social, and demographic information)?
- What symptoms has the patient reported? When did they begin, what has seemed to aggravate them, and what has provided relief?
- What physical signs have been noted on examination?
- How have signs and symptoms changed over time?
- What laboratory results have been, or are now, available?
- What radiologic and other special studies have been performed?
- What medications are being taken and are there any allergies?
- What other interventions have been undertaken?
- What is the reasoning behind the management decisions?

goal of answering three questions when the experiment is over:

1. What was the nature of the disease or symptom? 2. What was the treatment decision? 3. What was the outcome of that treatment? As is true for all experiments, one purpose is to learn from experience through careful observation and recording of data.

2.2.2 Support Communication Among Providers

A central function of structured data collection and recording in health care settings is to assist personnel in providing coordinated care to a patient over time.

It was once common for patients to receive essentially all their care from a single provider: the family doctor. In the world of modern medicine, the emergence of sub specialization and the increasing provision of care by teams of health professionals have placed new emphasis on the central role of the medical record in ensuring quality and continuity of care.

2.2.3 Anticipate Future Health Problems

Clinical data therefore are important in screening for risk factors, following patients' risk profiles over time, and providing a basis for specific patient education or preventive interventions, such as diet, medication, or exercise

2.2.4 Record Standard Preventive Measures

source of data on interventions that have been performed to prevent common or serious disorders. Sometimes the interventions involve counseling or educational programs immunizations

2.2.5 Identify Deviations from Expected Trends

Data often are useful in medical care only when viewed as part of a continuum over time. An example is the routine monitoring of children for normal growth and development by pediatricians. Single data points regarding height and weight may have limited use by themselves; it is the trend in such data points observed over months or years that may provide the first clue to a medical problem

2.2.6 Provide a Legal Record

The medical record is a legal document; the responsible individual must sign most of the clinical information that is recorded. a well-maintained record is a source of protection for both patients and their physicians.

2.2.7 Support Clinical Research

Medical knowledge also can be derived from the analysis of large patient data sets even when the patients were not specifically enrolled in an RCT, often referred to as **retrospective studies**.

2.3 Weaknesses of the Traditional Medical Record System

use of the traditional paper record is complicated and greatly limit the record's effectiveness

2.3.1 Pragmatic and Logistical Issues

Recall, first, that data cannot effectively serve the delivery of health care unless they are recorded

- The patient's paper chart may be unavailable when the health care professional needs may be in use by someone else at another location; it may have been misplaced despite the record-tracking system
- Once the chart is in hand, it might still be difficult to find the information required.
- difficult to read. It is not uncommon to hear one physician asking another as they peer together into a chart: "What is that word?"
- When a chart is unavailable, the health care professional still must provide patient care. Thus, providers make do without past data, basing their decisions instead on what the patient can tell them and on what their examination reveals.
- When patients who have chronic or frequent diseases are seen over months or years, their records grow so large that the charts must be broken up into multiple volumes. When a hospital clinic or emergency room orders the patient's chart, only the most recent volume typically is provided.

2.3.2 Redundancy and Inefficiency

- recording such information multiple times in different ways and in different locations within the chart, the combined bulk of these notes accelerates the physical growth of the document

2.3.3 Influence on Clinical Research

process is vulnerable to transcription errors.

n retrospective chart review to investigate a question that was not a subject of study at the time the data were collected and prospective studies in which the clinical hypothesis is known in advance and the research protocol is designed specifically to collect future data that are relevant to the question under consideration

difficulty in randomization & double blind which may impractical

extremely time-consuming

retrospective paper chart review is a laborious and tedious process and that people performing it are prone to make transcription errors and to overlook key data. One of the great appeals of EHRs (Chap. 12) is their ability to facilitate the chart review process. They obviate the need to retrieve hard copy charts; instead, researchers can use computer-based data retrieval and analysis techniques to do most of the work

2.3.4 The Passive Nature of Paper Records

A manual archival system is inherently passive; the charts sit waiting for something to be done with them. They are insensitive to the characteristics of the data recorded within their pages, such as legibility, accuracy, or implications for patient management. They cannot take an active role in responding appropriately to those implications.

2.4-2.5.2

Researchers are finding that the amount of data that they must manage and assess has become so large that they often find that they lack either the capabilities or expertise to handle the analytics that are required. This problem, sometimes dubbed the “big data” problem. These issues will undoubtedly influence the evolution of data systems and EHRs, as well as the growth of personalized medicine, in the years ahead.

Medicine is remarkable for its failure to develop a widely accepted standardized vocabulary and nomenclature.

people question whether it is possible to introduce too much standardization into a field that prides itself in humanism.

International Classification of Disease (ICD). The 10th revision of this standard, ICD10, is currently in use in much of the world.

Pathologists have developed another widely used diagnostic coding scheme; originally known as Systematized Nomenclature of Pathology (SNOP), it was expanded to the Systematized Nomenclature of Medicine (SNOMED).

- datum as a single observational point that characterizes a relationship.
- information refers to analyzed data that have been suitably curated and organized so that they have meaning.
- Knowledge, then, is derived through the formal or informal analysis (or interpretation) of information that was in turn derived from data. Thus, knowledge includes the results of formal studies and also common sense facts, assumptions, heuristics (strategic rules of thumb)

2.5.2 ...

The observation that patient Brown has a blood pressure of 180/110 is a datum, as is the report that the patient has had a myocardial infarction (heart attack). When researchers pool such data, creating information, subsequent analysis may determine that patients with high

blood pressure are more likely to have heart attacks than are patients with normal or low blood pressure. This analysis of organized data (information) has produced a piece of knowledge about the world. A physician's belief that prescribing dietary restriction of salt is unlikely to be effective in controlling high blood pressure in patients of low economic standing (because the latter are less likely to be able to afford special low-salt foods) is an additional personal piece of knowledge—a heuristic that guides physicians in their decision making.

Knowledge at one level of abstraction may be considered data at higher levels .

A database is a collection of individual observations without any summarizing analysis. An EHR system is thus primarily viewed as a data- base—the place where patient data are stored. When properly collated and pooled with other data, these elements in the EHR provide information about the patient. A knowledge base, on the other hand, is a collection of facts, heuristics, and models that can be used for problem solving and analysis of organized data (information). If the knowledge base provides sufficient structure, including semantic links among knowledge items, the computer itself may be able to apply that knowledge as an aid to case-based problem solving. Many decision-support systems have been called knowledge- based systems.

2.6 Strategies of Clinical Data Selection and Use

It is illusory to conceive of a “complete clinical data set.” All medical databases, and medical records, are necessarily incomplete because they reflect the selective collection and recording of data by the health care personnel responsible for the patient. Interpersonal differences in both style and problem solving that account for variations in the way practitioners collect and record data. Much of medical education is directed at helping physicians and other health professionals to learn what observations to make, how to make them (generally an issue of technique), how to interpret them, and how to decide whether they warrant formal recording.

An example of this phenomenon is the difference between the first medical history, physical examination by a medical student and a seasoned clinician examining the same patient. Medical students tend to work from comprehensive mental outlines of questions to ask, physical tests to perform, and additional data to collect. Because they have not developed skills of selectivity.

What do we mean by selectivity in data collection and recording? It is precisely this process that often is viewed as a central part of the “art of medicine,” an element that accounts for individual styles and the sometimes marked distinctions among clinicians.

The idea of selectivity implies an ongoing decision-making process that guides data collection and interpretation. Attempts to understand how expert clinicians internalize this process, and to formalize the ideas so that they can better be taught and explained, are central in biomedical informatics research. Improved guidelines for such decision making, derived from research activities in biomedical informatics.

2.6.1 The Hypothetico-Deductive Approach

Strategies for data collection and interpretation may be imbedded in an iterative process known as the hypothetico-deductive approach. The central idea is one of sequential, staged

data collection, followed by data interpretation and the generation of hypotheses, leading to hypothesis directed selection of the next most appropriate data to be collected. This process is iterated until one hypothesis reaches a threshold level of certainty (if you're still confused check the illustration on page 21 Fig. 2.12). At that point, a management, disposition, or therapeutic decision can be made. As medical students learn this process, their data collection becomes more focused and efficient, and their medical records become more compact.

Studies have shown that an experienced physician will have an initial set of hypotheses (theories) in mind after hearing the patient's response to the first six or seven questions.

Physicians refer to the set of active hypotheses as the differential diagnosis for a patient; the differential diagnosis comprises the set of possible diagnoses among which the physician must distinguish to determine how best to administer treatment. Note that the question selection process is inherently heuristic; e.g., it is personalized and efficient, but it is not guaranteed to collect every piece of information that might be pertinent. Human beings use heuristics all the time in their decision making because it often is impractical or impossible to use an exhaustive problem-solving approach. A common example of heuristic problem solving is the playing of a complex game such as chess.

Differences among such heuristics account in part for variations in observed expertise.

Physicians have developed safety measures to help them to avoid missing important issues that they might not discover when collecting data in a hypothesis-directed fashion when taking the history of a patient's present illness. These measures tend to be focused in four general categories of questions that follow the collection of information about the chief complaint: past medical history, family history, social history, and a brief review of systems in which the physician asks some general questions about the state of health of each of the major organ systems in the body. Occasionally, the physician discovers entirely new problems or finds important information that modifies the hypothesis list or modulates the treatment options available. When physicians have finished asking questions, the refined hypothesis list then serves as the basis for a focused physical examination. As in the question-asking process, focused hypothesis-directed examination is augmented with general tests that occasionally turn up new abnormalities and generate hypotheses that the physician did not expect on the basis of the medical history alone. Additional data gathering may still be necessary, however. Such testing is once again guided by the current hypotheses.

Ultimately, physicians are sufficiently certain about the source of a patient's problem to be able to develop a specific management plan. Data collected to measure response to treatment may themselves be used to synthesize information that affects the hypotheses about a patient's illness. If patients do not respond to treatment, it may mean that their disease is resistant to that therapy and that their physicians should try an alternate approach, or it may mean that the initial diagnosis was incorrect and that physicians should consider alternate explanations for the patient's problem. The patient may remain in a cycle of treatment and observation for a long time. This long cycle reflects the nature of chronic-disease management—an aspect of medical care that is accounting for an increasing proportion of the healthcare community's work (and an increasing proportion of health care cost).

2.6.2-2.7

sensitivity —the likelihood that a given datum will be observed in a patient with a given disease or condition

A pathognomonic observation is 100% specific for a given disease. Unfortunately, there are few pathognomonic tests in medicine and they are often of relatively low sensitivity

An observation is highly specific for a disease if it is generally not seen in patients who do not have that disease.

Yet even experienced physicians sometimes fail to recognize that, although they have made an observation that is highly specific for a given disease, it may still be more likely that the patient has other diseases (and does not have the suspected one) unless (1) the finding is pathognomonic or (2) the suspected disease is considerably more common than are the other diseases that can cause the observed abnormality.

baseline rate: the prevalence in the population from which the patient was selected

predictive value (PV) of a test: is simply the post-test (updated) probability that a disease is present based on the results of a test.

If an observation supports the presence of a disease, the PV will be greater than the prevalence (also called the pretest risk). If the observation tends to argue against the presence of a disease, the PV will be lower than the prevalence. For any test and disease, then, there is one PV if the test result is positive and another PV if the test result is negative. These values are typically abbreviated PV+ (the PV of a positive test) and PV- (the PV of a negative test).

The PV+ formula is one of many forms of Bayes' theorem