Theme 2. Medical data in computers. Coding and classification of medical information.

Information can be accessed in different forms and by different types of media and carriers. The different types of carriers that transport information are:

1. ***Integers:*** a discrete number (e.g., the number of cases in a population: this population can be human or a sample of specimen or age in years, months or number of hospitals or health units);

2. ***Reals***: a measured variable (e.g., a temperature or a blood pressure);

3. ***Codes***: an observation (e.g., pain or a swelling or codes representing diseases or drugs);

4. ***Text:*** natural language (e.g., text in the case reporting giving explanation of things or annotation in the form of events, description, etc.);

5. ***Graphics and maps*** in the form of Geographic Information Systems, diagrams, x-Ray images, charts.

A most important issue related to the recording of medical data in a computer and its presentation is the ***completeness, reliability*** and ***precision.*** These factors together make the quality of data regardless of its type:

1. **Completeness**. Incomplete data may result in uncertainty. In case reporting it is not always clear whether data are missing or are absent because they were considered to be irrelevant, or were just not documented in the data collection sheet;

2. **Accuracy** is the ability to perform a task without making mistakes or errors, or it is the degree of conformity of a measure to a certain standard or a true value. The former meaning can be characterized as correctness; the second one can be characterized as conformity or exactness. **Correctness** is a measure of the error rate of the data. Errors are first made during data collection, either in observations or in measurements. The observed value also has a deviation due to reading or measuring errors. We have here to distinguish between the systematic and the statistical errors. **Conformity** of data pertains to following standards or classification systems for data recording. When classification and coding systems are used to document cases, we need to follow the rules and use the definitions of the classification system to select a proper code;

3. **Precision** deals with the degree of refinement or granularity by which a measurement is expressed, such as the number of decimal places. A body weight expressed as 89.12 kg expresses a higher precision than a weight expressed as 89.1 kg. It is misleading to specify a value with a higher precision than the accuracy with which the value is obtained. It is equally misleading that the weight of an adult for example should be expressed with two decimals if this precision does not bear any meaning;

4. **Coding.** A factor related to the quality of data is coding. In **coding** data, the user should first interpret the data and then assign a code. Interpretation errors are inherent to coding. On the one hand, the coding of data limits the way of expressing oneself, but on the other hand, it enforces **standardization of terminology**, which is extremely important if the data is to be used by people other than a data collector, which is the case in all health indicators data;

Text, images, sounds, and many other sources of data can be stored and retrieved by computers, but for computers to “understand” the data it must be put into ***a code***. To manage patients within an integrated health service it is becoming more important to communicate information from one setting in a digital format while retaining meaning when viewed in a different context.

All of our lives have been affected in one way or another by the use of coded medical information. For people who are seeking health insurance coverage, medical coding is used to determine if a person is a "good" risk for health insurance coverage. A mistake in medical coding could result in denial of coverage. When it comes to patient care, coding determines which

health care services are reimbursed and how much is paid. If an error is made in the coding diagnosis or treatment, payment could be denied or too much can end up being paid for services.

In conclusion, several important rules can be phrased, based on the quality factors that have been emphasized above, to enable better recording of medical data in computers:

- Data should be acquired as close to the source of the data as possible;
- Data should be recorded by obeying strict rules of standardization and coding;
- The original data should be stored, and if possible, human interpretations should be stored only if the raw data they are based on are also stored;
- Coding of data should be done only if there is no other way to present the data, and it should preferably be done by the person making the observation;
- For all data entered, there should preferably be an on-line **feedback** to the user to signal possible deviations from what should be expected;
- Persons who enter the data should ideally benefit from this data entry, either because they will use the data later on or because it will improve the quality of their work;
- Authentication of data (adding the coder's name and signature) and time stamping of data improves the data quality.

Data processing refers to the process of performing specific operations on a set of data or a database.

Data processing primarily is performed on information systems, a broad concept that encompasses computer systems and related devices. At its core, an information system consists of input, processing, and output. In addition, an information system provides for feedback from output to input. The input mechanism (such as a keyboard, scanner, microphone, or camera) gathers and captures raw data and can be either manual or automated. Processing, which also can be accomplished manually or automatically, involves transforming the data into useful outputs. This can involve making comparisons, taking alternative actions, and storing data for future use. Output typically takes the form of reports and documents that are used by managers. Feedback is utilized to make necessary adjustments to the input and processing stages of the information system.

The processing stage is where management typically exerts the greatest control over data. It also is the point at which management can derive the most value from data, assuming that powerful processing tools are available to obtain the intended results. The most frequent processing procedures available to management are basic activities such as segregating numbers into relevant groups, aggregating them, taking ratios, plotting, and making tables. The goal of these processing activities is to turn a vast collection of facts into meaningful nuggets of information that can then be used for informed decision making, corporate strategy, and other managerial functions.

The traditional approach to data management consists of maintaining separate data files for each application. One or more data files are created for each application. However, duplicated files results in data redundancy. The problem with data redundancy is the possibility that updates are accomplished in one file but not in another, resulting in a lack of data integrity. Likewise, maintaining separate files is generally inefficient because the work of updating and managing the files is duplicated for each separate file that exists. To overcome potential problems with traditional data management, the database approach was developed.

Classification is a method for systematically grouping something - for example, diseases. In most classifications, classes are designated by codes, which allow aspects of the things to be captured (a systematic arrangement of similar entities on the basis of certain characteristics).

A code is usually a unique numeric or alphabetic representation of items in a classification.

Nomenclature is a system of naming used in a branch of knowledge. Medical nomenclature attempts to standardize the names used for patient findings, diseases, interventions, and outcomes.

Medical classification and coding is a standards-based translation of diagnoses and procedures. Its original purpose was to retrieve health record data for research. However, over time, healthcare professionals recognized the applied value towards patient care since coding and classification converts long narrative descriptions of diagnoses and procedures into terminologies that can be used to render healthcare services. Medical coding also provides tangible benefits in the administration of healthcare.

Coding and classification is used by various organizations for the tracking and reporting of diseases (epidemiology), for medical research (necessitating interoperable vocabularies), and to assess the quality and cost of healthcare services.

Classification and coding is important in helping to explain the dynamic linkages between diseases and their symptoms. As a consequence, classification and coding systems must be adaptable and flexible. Indeed, effective coding for clinical disease classification is a crucial part of on-going efforts aimed at improving the quality of healthcare.

History of medical classification

In 1893, the International Statistical Institute approved a standardized system for classifying deaths. The list was prepared by a Paris statistician, Jacque Bertillon, and was called the Bertillon Classification. By 1900, 26 countries had implemented the Bertillon Classification. In 1928, a study sponsored by the Health Organization of the League of Nation discussed how the Bertillon Classification could be expanded to include the tracking of diseases.

In 1949, the World Health Organization (WHO) realized the idea of enacting a system for tracking mortality as well as causes of diseases on a global basis. The Manual of the International Classification of Diseases, Injuries and Causes of Death (ICD) has served as the foundation for the modern practice of medical coding. Starting with the Bertillon Classification, this list is revised every 10 years. The latest revision is ICD-10.

Medical classification systems are used for a variety of applications in medicine and medical informatics

- Statistical analysis of diseases and therapeutic actions
- Knowledge-based and decision support systems
- Direct surveillance of epidemic or pandemic outbreaks

There are country specific standards and international classification systems.

Types of coding systems specific to health care include:

1. ***Diagnostic codes***
   - are used to determine diseases, disorders, and symptoms;
   - can be used to measure morbidity and mortality;
   - examples: ICD-9-CM, ICD-10.
2. ***Procedural codes***
   - they are numbers or alphanumeric codes used to identify specific health interventions taken by medical professionals;.
   - examples: ICPM, ICHI.
3. ***Pharmaceutical codes***
   - are used to identify medications;
   - examples: ATC, NDC
4. ***Topographical codes***
   - are codes that indicate a specific location in the body
   - examples: ICD-O, SNOMED.

The World Health Organization (WHO) maintains several internationally endorsed classifications designed to facilitate the comparison of health related data within and across populations and over time as well as the compilation of nationally consistent data. This **"Family of International Classifications" (FIC)** include three main (or reference) classifications on basic parameters of health prepared by the organization and approved by the World Health Assembly for international use, as well as a number of derived and related classifications providing additional details. Some of these international standards have been revised and adapted by various countries for national use.

1. **Reference classifications**

International Statistical Classification of Diseases and Related Health Problems (ICD)

ICD-10 (10th revision, in use by WHO since 1994)

ICD-10-CM (Clinical Modification, used in the US since October 2015)

ICD-10-CA (used for morbidity classification in Canada)

ICD-10-AM (used in Australia and New Zealand)

International Classification of Functioning, Disability and Health (ICF)

International Classification of Health Interventions (ICHI)

2. **Derived classifications**

International Classification of Diseases for Oncology, Third Edition (ICD-O-3)

ICD-10 Classification of Mental and Behavioural Disorders

Application of the International Classification of Diseases to Dentistry and Stomatology, 3rd Edition (ICD-DA)

Application of the International Classification of Diseases to Neurology (ICD-10-NA)

EUROCAT is an extension of the ICD-10 Q chapter for congenital disorders

3. **Related classifications**

International Classification of Primary Care (ICPC)

ICPC-2 PLUS

International Classification of External Causes of Injury (ICECI)

Anatomical Therapeutic Chemical Classification System with Defined Daily Doses (ATC/DDD)

Technical aids for persons with disabilities: Classification and terminology (ISO9999)

International Classification for Nursing Practice (ICNP)

The International Statistical Classification of Diseases and Related Health Problems were developed by the WHO to harmonize disease diagnosis criteria.

The ICD has become the international standard diagnostic classification. Its scope includes the analysis of the health situation of various population groups and the monitoring of the incidence and prevalence of diseases and other health problems in relation to other variables. Furthermore, the ICD can be used to classify diseases and other health problems recorded on many types of health and vital records, including death certificates and other hospital records. In addition, the ICD allows for the storage and search for diagnostic information intended for clinical and epidemiological purposes. Health records provide the basis for the WHO Member States to generate national mortality and incidence rates. Moreover, the Classification includes the proper format of death cause reporting in death certificates.

Nowadays, the 10th edition of the Classification [ICD-10] is available, which includes:

- tabular lists of death cause names and codes;
- terms for death cause names, including the ones excluded from a given category or included in it;
- alphabetical annex of diseases and natural injuries, external causes of injuries, as well as tables of medicines and chemical substances;
- description, guidelines, and rules for encoding.

The International Classification of Medical Procedures, Revision 9 CM is a classification of medical procedures fulfilled by healthcare facilities [primarily hospitals], with the main focus on treatment, diagnosis and therapy. The ICD-9-CM includes:

- tabular index of numerical disease codes
- alphabetical index of diseases
- classification system for operational, diagnostic, and therapeutic procedures [alphabetical index and tabular index].

The Systematized Nomenclature of Medicine [SNOMED] has been developed by the College of American Pathologists [CAP]. It is a multiaxial, hierarchical classification system. As in any such system, a disease may be located in a body organ which results in a code in a topography axis and may lead to morphological alterations represented by a morphology code. It systematically organizes computer processable collection of medical terminology covering most areas of clinical information such as diseases, findings, procedures, microorganisms, pharmaceuticals etc. It allows a consistent way to index, store, retrieve, and aggregate clinical data across specialties and sites of care.

SNOMED is a compositional concept system, which means that concepts can be specialised by combinations with other concepts.

The SNOMED encoding system enables automatic data processing without information loss.

The structure of SNOMED is as follows:

- topography [T] – specifies body site, organ or part
- morphology [M] – specifies pathological changes in organs, tissues or cells
- etiology [E] – specifies cause, e.g. injuries
- function [F]
- disease [D]
- procedure [P]
- occupancy [J].

SNOMED CT (Systematized Nomenclature of Medicine -- Clinical Terms) is a standardized, multilingual vocabulary of clinical terminology that is used by physicians and other health care providers for the electronic exchange of clinical health information.

SNOMED CT currently contains more than 300,000 medical concepts, divided into hierarchies as diverse as body structure, clinical findings, geographic location and pharmaceutical/biological product. Each concept is represented by an individual number and several concepts can be used simultaneously to describe a complex condition.

By using numbers to represent medical concepts, SNOMED CT provides a standard by which medical conditions and symptoms can be referred, eliminating the confusion that may result from the use of regional or colloquial terms. The numerical reference system also facilitates the exchange of clinical information among disparate health care providers and electronic medical records (EMR) systems.

Health Level Seven (HL7) is a data standard for communication/messages between:

- Patient administrative systems (PAS)
- Electronic practice management
- Lab information systems (interfaces)
- Dietary
- Pharmacy (clinical decision support)
- Billing
- Electronic health records

Digital Imaging and Communications in Medicine (DICOM) is the standard for the communication and management of medical imaging information and related data. It was formed by the National Electrical Manufacturers Association (NEMA) and the American College of Radiology.

The DICOM Standard facilitates interoperability of medical imaging equipment by specifying:

- For network communications, a set of protocols to be followed by devices claiming conformance to the Standard.
- The syntax and semantics of Commands and associated information that can be exchanged using these protocols.
- For media communication, a set of media storage services to be followed by devices claiming conformance to the Standard, as well as a File Format and a medical directory structure to facilitate access to the images and related information stored on interchange media.
- Information that must be supplied with an implementation for which conformance to the Standard is claimed.

This Standard has been developed with an emphasis on diagnostic medical imaging as practiced in radiology, cardiology, pathology, dentistry, ophthalmology and related disciplines, and image-based therapies such as interventional radiology, radiotherapy and surgery. However, it is also applicable to a wide range of image and non-image related information exchanged in clinical, research, and other medical environments.

Control questions

1. What types of carriers that transport medical information are known?
2. What is a most important issue related to the recording of medical data in a computer and its presentation?
3. What are important rules used for recording of medical data in computers?
4. What is data processing?
5. What are classifiction, code, nomenclature?
6. What is medical classification and coding?
7. Why is medical classification and coding important?
8. Types of medical classification systems.
9. What are SNOMED, HL7, DICOM?

References

1. Basic concepts in medical informatics J. Epidemiol. Community Health. 2002;56(11):808-812.
2. Biomedical Informatics in the Education of Physicians JAMA. 2010;304(11):1227-1228.
3. Biomedical Informatics: Computer Applications in Health Care and Biomedicine JAMA. 2006;296(21):2624-2625.
4. Dr. Najeeb Al-Shorbaji.Health and Medical Informatics:Technical Paper By, RA/HIS Health Information Support Regional Office for the Eastern Mediterranean World Health Organization. - Cairo, Egypt. - May 2001. -18 p.
5. Handbook of Medical Informatics /J.H. van Bemmel, Erasmus University, Rotterdam, M.A. Musen, Stanford University, Stanford http://www.mihandbook.stanford.edu/handbook/home.htm. Accessed 15 may 2005.
6. Aalseth, P. (2006). Medical Coding. Sudbury: Jones and Bartlett.
7. Shortliffe, E., Cimino, J. (2006). Biomedical Informatics. New York: Springer.
8. Starren, J., Johnson, S. (1999). An Object Oriented Taxonomy of Medical Data