## Healthcare Data Sources and Basic Analytics:-

This part discusses the details of various healthcare data sources and the basic analytical methods that are widely used in the processing and analysis of such data. The various forms of patient data that is currently being collected in both clinical and non-clinical environments will be studied.

The clinical data will have the structured electronic health records and biomedical images. Sensor data has been receiving a lot attention recently.

Personalized medicine has gained a lot of importance due to the advancements in genomic data.

# Data Source

## e.g. Data

| | Non-Sensitive Data | | | Sensitive Data | |
|---|---|---|---|---|---|
| # | Zip | Age | Nationality | Name | Condition |
| 1 | 13053 | 28 | Indian | Kumar | Heart Disease |
| 2 | 13067 | 29 | American | Bob | Heart Disease |
| 3 | 13053 | 35 | Canadian | Ivan | Viral Infection |
| 4 | 13067 | 36 | Japanese | Umeko | Cancer |

Identity data - e.g., name, address, personal number, etc.

Demographic data - e.g., sex, age, nationality, etc.

Analysis data - e.g., diseases, habits, salary, average income, crimes, sales etc.

Quasi-identifier : age, gender , zip…

Sensitive data : salary, disease …

Non-sensitive : others

- Healthcare data is particularly rich and it is derived from a wide variety of sources such as **sensors, images, text**.
- It is in the form of **biomedical literature/clinical notes**, and traditional electronic records.
- This heterogeneity in the **data collection and representation process** leads to numerous challenges.
- There is a **wide diversity in the techniques that are required to analyze these different forms of data.**
- In addition, the **heterogeneity of the data naturally creates various data integration and data analysis challenges**.
- In many cases, **insights can be obtained from diverse data types, which are otherwise not possible from a single source of the data.**
- It is only **recently that the vast potential of such integrated data analysis methods is being realized.**

# Example

**Hospital A**

Patient Name: Vivian Christensen
Visit ID: 837720
Date of Birth: 3/20/1953
SSN: 000-86-6628
MRN: 9968427
Dx: 414.00

**Hospital B**

Patient Name: Viv Christensen
Visit ID: 483005
Date of Birth: 3/20/1953
SSN: 000-68-6628
MRN: 0099523461
Dx: 493.01

## CONTD..

- Often **data from different health care sources** need to be **collected and analyzed**

  1. Improve data quality

  II. Enrich data with additional information

  III. Allow data analyses that are impossible on individual databases

- **Can this data be utilized?**
  - For improving patient treatment and care
  - For various health care medication
  - For medical research

# Healthcare Data Sources and Basic Analytics

## 1. Electronic Health Records

- **Electronic health records (EHRs)** contain a **digitized version of a patient's medical history**.
- It encompasses a **full range of data relevant to a patient's care** such as demographics, problems, medications, physician's observations, vital signs, medical history, laboratory data, radiology reports
- Many EHRs go beyond a **patient's medical or treatment history and may contain additional broader perspectives of a patient's care.**
- An **important property of EHRs** is that they provide an **effective and efficient way for healthcare providers and organizations to share with one another.**
- In this context, EHRs are inherently **designed to be in real time and they can instantly be accessed and edited by authorized users.**
- This can be very useful in practical settings. For example, **a hospital or specialist may wish to access the medical records of the primary provider.**
- It improve **quality and convenience of patient care**, **increase patient participation in the healthcare process**, **improve accuracy of diagnoses and health outcomes**, and i**mprove care coordination.**
- Various components of EHRs along with the advantages, barriers, and challenges of using EHRs
- are discussed in further section.

## 2. **Biomedical Image Analysis**

- Medical imaging plays an important role due to its immense capability in providing **high-quality images of anatomical structures** in human beings.

- The most popular imaging modalities used to acquire a biomedical image are **magnetic resonance imaging (MRI)**, **computed tomography (CT), positron emission tomography (PET),** and **ultrasound (U/S)**.

- Being **able to look inside of the body without hurting the patient** and being **able to view the human organs has tremendous implications** on human health.

- However, merely viewing such organs with the help of images is just the first step of the process. The final goal of **biomedical image analysis is to be able to generate quantitative information and make inferences from the images** that can provide far more insights into a medical condition.

- Such **analysis has major societal significance**. However, it includes **many challenges since the images are varied, complex, and can contain irregular shapes with noisy values.**

- A number of **general categories of research problems that arise in analyzing images** are **object detection**, **image segmentation**, **image registration**, and **feature extraction**.

Fig. anatomical imaging

Fig. MRI


Fig. CT SCAN


Fig. PET


Fig. ultrasound (U/S)

## 3 **Sensor Data Analysis**

- Sensor data is ubiquitous in the medical domain both for real time and for retrospective analysis.

- Several **forms of medical data collection instruments such as ECG,EEG** are essentially **sensors that collect signals from various parts of the human body**. These collected data instruments are sometimes used for real-time analysis.

- Perhaps, the most important **use-case of real-time analysis is in the context of ICUs**.
- In all these cases, the volume of the data to the processed can be rather large.
- **For example**, **in an ICU**, it is **not uncommon for the sensor to receive input from hundreds of data sources, and alarms need to be triggered in real time.** Such applications necessitate the use of big-data frameworks and specialized hardware platforms.

4. **Biomedical Signal Analysis**

- Consists of **measuring signals from biological sources**. Examples: **ECG, EEG** and so on**.**
- The **analysis is vital in diagnosing the pathological conditions** and in deciding an appropriate care pathway.
- The measurement of physiological signals gives some form of quantitative or relative assessment of the state of the human body.
- These **signals are acquired from various kinds of sensors and transducers** ( discrete or continuous).
- The **processing and interpretation of physiological signals is challenging due to the low signal-to-noise ratio (SNR)** and the interdependency of the physiological systems.
- Require a **significant amount of preprocessing**.
- Principal Component Analysis (PCA), Singular Value Decomposition (SVD).

## 5 Genomic Data Analysis

- A **significant no. of diseases are genetic in nature**, but the nature of the causality between the genetic markers and the diseases has not been fully established. e.g. **Diabetes**

- Mostly interested in understanding what kind of health-related questions can be addressed through in-silico analysis of the genomic data. Moreover, translating genetic discoveries into personalized medicine practice is a highly non-trivial task.

- For example, **the genomic complex diseases such as cancers are overwhelmingly complicated, revealing a high order of heterogeneity among different individuals.**



- **Genetic data** are often **modeled either as sequences or as networks.** Therefore, the work in this field requires a good understanding of sequence and network mining techniques.

- Various data analytics-based **solutions are being developed for tackling key research problems in medicine such as identification of disease biomarkers** and prediction of clinical outcome.

# 6. Clinical Text Mining

Most of the information about patients is encoded in clinical notes. These notes are typically stored in an unstructured data format.

These contain the clinical information from the transcription of dictations, direct entry by providers, or use of speech recognition applications.

The manual encoding of this free-text form on a broad range of clinical information is too costly and time consuming.

It becomes hard mainly because of their unstructured nature, heterogeneity, diverse formats, and varying context across different patients and practitioners.

NLP and entity extraction play an important part in inferring useful knowledge from large volumes of clinical text.

The processing of clinical text using NLP methods is more challenging due to the ungrammatical nature of short and telegraphic phrases, dictations, shorthand lexicons such as abbreviations and acronyms, and often misspelled clinical terms.

# 7 Mining Biomedical Literature

- A significant number of applications rely on evidence from the **biomedical literature.**
- The **use of text mining methods** for the **long-term preservation, accessibility, and usability of digitally available resources is important in biomedical applications** relying on evidence from scientific literature.
- **Text mining methods and tools offer novel ways of applying new knowledge discovery** methods in the biomedical field.
- One of the major **challenges in biomedical text mining is the multidisciplinary nature of the field**. For example, **biologists describe chemical compounds using brand names, while chemists often use less ambiguous IUPAC-compliant names or unambiguous descriptors such as International Chemical Identifiers.**
- While the latter can be handled with **various chem-informatics tools , text mining techniques are required to extract less precisely defined entities and their relations** from the literature.
- In this context, **entity and event extraction methods play a key role in discovering useful knowledge from unstructured databases.**
- **Text mining** brings about other benefits to **biomedical research by linking textual evidence to biomedical pathways**, **reducing the cost of expert knowledge, validation, and generating hypotheses.**

## 8 **Social Media Analysis**

- The **rapid emergence of various social media resources** such as social networking sites, blogs/microblogs, forums, question answering services, and online communities **provides a wealth of information about public opinion on various aspects of healthcare.**

- **Social media data can be mined for patterns and knowledge** that can be leveraged to make useful inferences about population health and public health monitoring.

- A significant amount of **public health information can be gleaned from the inputs of various participants at social media sites.**

- Although **most individual social media posts and messages contain little informational value, aggregation of millions of such messages can generate important knowledge** .Effectively analyzing these vast pieces of knowledge can significantly reduce the latency in collecting such complex information.
- Previous research on social media **analytics for healthcare has focused on capturing aggregate health** trends such as **outbreaks of infectious diseases, detecting reports of adverse drug interactions, and improving interventional capabilities for health-related activities.**

- Disease outbreak detection is often strongly reflected in the content of social media and an analysis of the history of the content provides valuable insights about disease outbreaks.

- Topic models are frequently used for high-level analysis of such health-related content.

# Applications and Practical Systems for Healthcare

## 1. Data Analytics for Pervasive Health

Pervasive health refers to the **process of tracking medical well-being and providing long-term medical care** with the use of advanced technologies such as **wearable sensors**. For example, **wearable monitors are often used for measuring the long-term effectiveness of various treatment mechanisms**.

These methods, however, face a number of **challenges, such as knowledge extraction from the large volumes of data collected and real-time processing**.

However, **recent advances** in both **hardware and software technologies (data analytics in particular**) have made such systems a reality.

In the case of wearable sensors, **sensors are attached to the body or woven into garments**. For example, **3-axis accelerometers distributed over an individual's body can provide information about the orientation and movement of the corresponding body part.**

## 2 Healthcare Fraud Detection

- With growing healthcare costs, the threat of healthcare fraud is increasing at an alarming pace.

- Analyze the healthcare claims data along different dimensions to identify fraud.

- The complexity of the healthcare domain, which includes multiple sets of participants, makes the problem of detecting healthcare fraud equally challenging and makes it different from other domains such as credit card fraud detection and auto insurance fraud detection.

- The users in the healthcare setting are the beneficiaries, who typically are not the fraud perpetrators. Hence, more sophisticated analysis is required in the healthcare sector to identify fraud.

# 3. **Data Analytics for Pharmaceutical Discoveries**

- The cost of **successful novel chemistry-based drug development** often reaches **millions of dollars, and the time to introduce the drug to market** often comes close to a decade.
- The **high failure rate of drugs during this process, make the trial phases known as the "valley of death**." Most **new compounds fail during the FDA approval process in clinical trials or cause adverse side effects.**

- In the context of pharmaceutical discoveries, **data analytics can potentially limit the search space and provide recommendations to the domain experts** for hypothesis generation and further analysis and experiments.

- **Data analytics** can be used in **several stages of drug discovery and development to achieve different goals.** In this domain, one way to categorize data analytical approaches is based on their application to pre-marketing and post-marketing stages of the drug discovery and development process.
- In the pre-marketing stage, data analytics focus on discovery activities such as finding **signals that indicate relations between drugs and targets**, **drugs and drugs, genes and diseases, protein and diseases, and finding biomarkers.**

- In the post-marketing stage an **important application of data analytics is to find indications of adverse side effects for approved drugs**.

# 4 Clinical Decision Support Systems

- **Clinical Decision Support Systems (CDSS)** are computer systems designed to assist clinicians with **patient-related decision making, such as diagnosis and treatment** .

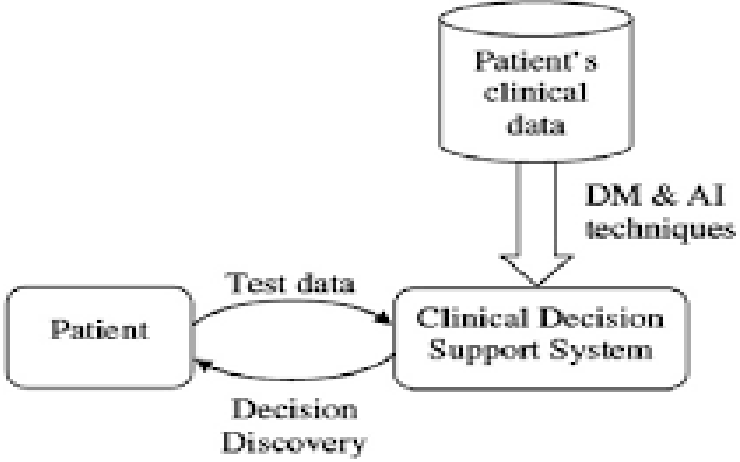
- CDSS have become a **crucial component in the evaluation and improvement of patient treatment since they have shown to improve both patient outcomes and cost of care**.

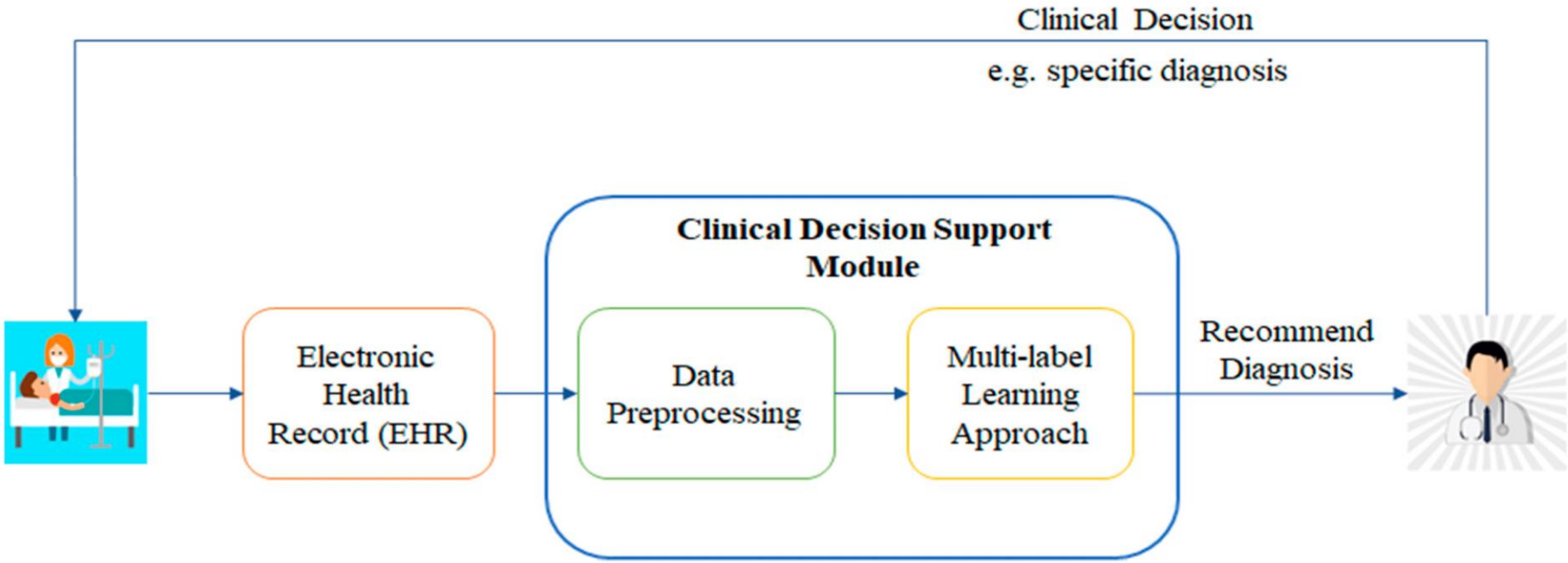
- They can help in **minimizing analytical errors by notifying the physician of potentially harmful drug interactions, and their diagnostic procedures have been shown to enable more accurate diagnoses.**

- Some of the main advantages of **CDSS are their ability in decision making and determining optimal treatment strategies, aiding general health policies by estimating the clinical and economic outcomes of different treatment methods under certain conditions**.

- The main **reason for the success of CDSS are their electronic nature, seemless integration with clinical workflows, providing decision support at the appropriate time/location.**

- Two **particular fields of healthcare where CDSS** have been extremely influential are **pharmacy and billing.** CDSS can help pharmacies to look for negative drug interactions and then report them to the corresponding patient's ordering professional.


Patient's clinical data
DM & AI techniques
Test data
Patient
Clinical Decision Support System
Decision Discovery


Clinical Decision
e.g. specific diagnosis
Clinical Decision Support Module
Electronic Health Record (EHR)
Data Preprocessing
Multi-label Learning Approach
Recommend Diagnosis

## 5 Computer-Aided Diagnosis

- **Computer-aided diagnosis/detection (CAD)** is a **procedure in radiology that supports radiologists in reading medical images.**

- CAD tools can improve the **performance of the radiologist**. CAD algorithms is running in the background or **has already been precomputed. Structures identified by the CAD algorithm are then highlighted as regions of interest to the radiologist**.

- The principal value of CAD tools is **determined not by its stand-alone performance, but rather by carefully measuring the incremental value of CAD in normal clinical practice,** such as the number of additional lesions detected using CAD.

- Secondly, **CAD systems must not have a negative impact on patient management** (for instance, false positives that cause the radiologist to recommend unnecessary biopsies and followups).

- From the data analytics perspective, new **CAD algorithms aim at extracting key quantitative features, summarizing vast volumes of data, and/or enhancing the visualization of potentially malignant nodules, tumors, or lesions in medical images**.

# 6. Mobile Imaging for Biomedical Applications

- Mobile imaging refers to the **application of portable computers such as smartphones or tablet computers to store, visualize, and process images** with and without connections to servers, the Internet, or the cloud.

- Today, **portable devices provide sufficient computational power for biomedical image processing and smart devices** have been introduced in the operation theater.

- While many techniques for **biomedical image acquisition will always require special equipment, the regular camera is one of the most widely used imaging modality in hospitals.** Mobile technology and smart devices, especially smartphones, **allows new ways of easier imaging**.

# Before the use of Electonic Health Record

# Before the use of Electronic Health Record

# Before the use of Electonic Health Record

# Need for EHR (CDSS)

- **Medical errors are costly**
  - **Charges/Costs**
  - **Morbidity/Mortality**
- **CDSS technology can help reduce**
  - **errors**
  - **costs**
- **EHR**
  - **Collection and organization of data**
  - **Vehicle for decision support**

# Developing Standards for EHR Birth and Death Data Exchange with Vital Records Systems

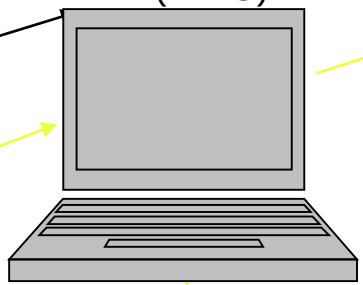

Mother

Mother's Worksheet

Birth Registration System (EBRS)


Registrar


Birth Information Specialist




Birth Certificate

State Department of Health


Birth Event


Electronic Health Record


CDC/NCHS


Nurse


Physician

# Objectives

- Discuss the core elements of a Master Patient Index (MPI)

- Describe the components integral to a Unique Patient Identifier (UPI)

# Patient Management and Billing

- Systems that support patient management functions
  - Example
    - Patient identification
      - Master patient index

# Master Patient Index (MPI)

- An index referencing all patients known to an area, enterprise or organization
- Synonyms
  - Patient Master Index (PMI)
  - Master Person Index
  - Enterprise Master Patient Index (EMPI)

-

- List or database created and maintained by a healthcare facility to record the name and identification number of every patient who has ever been admitted or treated in the facility

# Purpose of the MPI

- Provides the index, location of, and access to a patient's EHR in an enterprise.
- Facilitates intraoperability and the accurate creation of a longitudinal record.
- Ensures accurate and complete linking of EHRs for HIE

# Purpose of the MPI

- Establishes a streamlined governance process
- Accurately matches persons being registered for care with their existing medical records
- Reconciling and Managing Systems

# Core Elements of a MPI

- Person name
- Alias/previous/maiden names
- Date of birth
- Address
- Telephone number
- Health record number

# Unique Patient Identifier (UPI)

- Value permanently assigned to an individual for identification purposes
- Unique across the entire national healthcare system
- Not shared with any other individual

# Functions a UPI Must Support

- Identification of an individual
- Identification of information
- Accurate identification functions
- Reduce healthcare operational cost and enhance the health status of the nation

# Components Integral to UPI

- An Identifier scheme
- Identification information
  - Permanent data segment
  - Longitudinal data segment
  - Health service data segment

# Components Integral to UPI

- Index
  - Organizational MPI
  - Enterprise-wide MPI
  - Registry MPI
- Information from previous episodes of care and different sites of care

# Components Integral to UPI

- Mechanism to hide or encrypt the Identifier
- Technology infrastructure to search, identify, match, encrypt, etc.
- Administrative infrastructure including the Central Governing Authority

# Processes Integral to Patient Identification

- Scope of access
  - Within a single organization
  - Enterprise wide access
  - Nation wide access

## 1.2 **Electronic Health Records(EHR)**

- Components of EHR,
- Benefits of EHR,
- Barriers to Adopting EHR,
- Challenges of using EHR data,
- Phenotyping Algorithms

## Electronic Health Records(EHR)

- Electronic health records (EHRs) contain a **digitized version of a patient's medical history**. It encompasses a full range of data relevant to a **patient's care such as demographics, problems, medications, physician's observations, vital signs, medical history, laboratory data, radiology reports, progress notes, and billing data**.
- Many EHRs go **beyond a patient's medical or treatment history and may contain additional broader perspectives of a patient's care**. An important property of EHRs is that they provide an **effective and efficient way for healthcare providers and organizations to share with one another**. In this context, EHRs are inherently **designed to be in real time and they can instantly be accessed and edited by authorized users.**
- This can be very useful in practical settings. For example, **a hospital or specialist may wish to access the medical records of the primary provider. An electronic health record streamlines the workflow by allowing direct access to the updated records in real time [30].**
- It **can generate a complete record of a patient's clinical encounter, and support other care-related activities such as evidence-based decision support, quality management, and outcomes reporting.**
- The **storage and retrieval** of health-related data is more efficient using EHRs. It helps to **improve quality and convenience of patient care, increase patient participation in the healthcare process, improve accuracy of diagnoses and health outcomes, and improve care coordination** [29].

- The main purpose of EHR is to **support clinical care and billing**.

- This also includes other functionalities, such as **improving the quality and convenience of patient care, improving the accuracy of diagnoses and health outcomes, improving care coordination and patient participation, improving cost savings, and finally, improving the general health of the population.**
- Most modern EHR systems are designed **to integrate data from different components such as administrative, nursing, pharmacy, laboratory, radiology, and physician' entries, etc.**
- **Electronic records may be generated from any department**.
- Hospitals and clinics may **have a number of different ancillary system providers**; in that case, these **systems are not necessarily integrated to the main EHR system.**
- It is possible that these **systems are stand-alone, and different standards of vocabularies have been used.**
- If **appropriate interfaces are provided, data from these systems can be incorporated in a consolidated fashion**; **otherwise a clinician has to open and log into a series of applications to get the complete patient record**.
- The **number of components present** may also **vary depending on the service provided**. Figure 2.1 shows different components of an EHR system

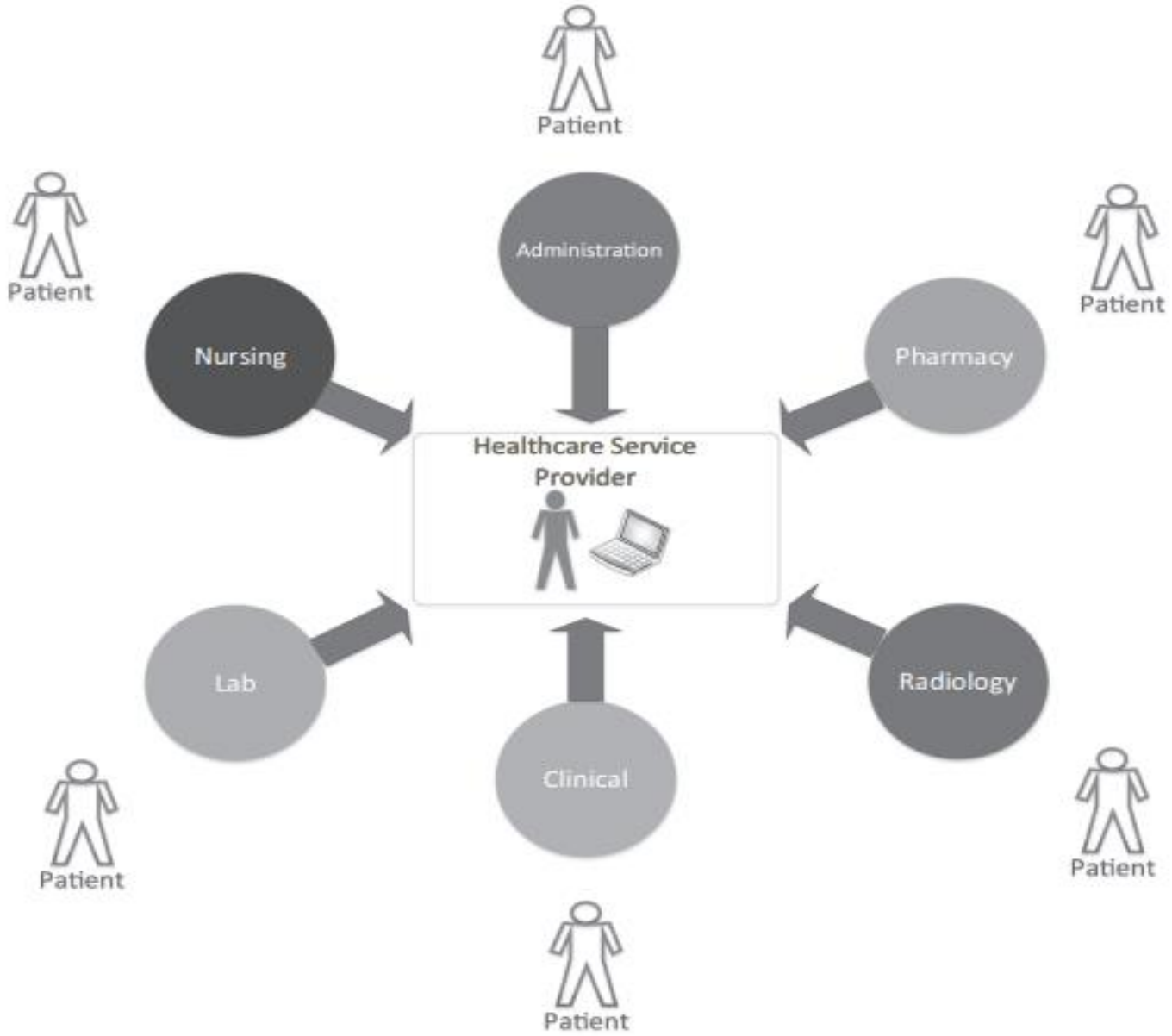



**FIGURE 2.1**: Various components of EHR.

# Developing Standards for EHR Birth and Death Data Exchange with Vital Records Systems


Mother

Mother's Worksheet


Birth Information Specialist

Birth Registration System (EBRS)


Registrar


Birth Certificate

State Department of Health


Birth Event


Electronic Health Record


Nurse


Physician


CDC/NCHS

## Various components of EHRs

1. Administrative System Components . . . . . .
2. Laboratory System Components & Vital Signs . . .
3. Radiology System Components . . . . . .
4. Pharmacy System Components . . . . . .
5. Computerized Physician Order Entry (CPOE) . .
6. Clinical Documentation . . . . . .

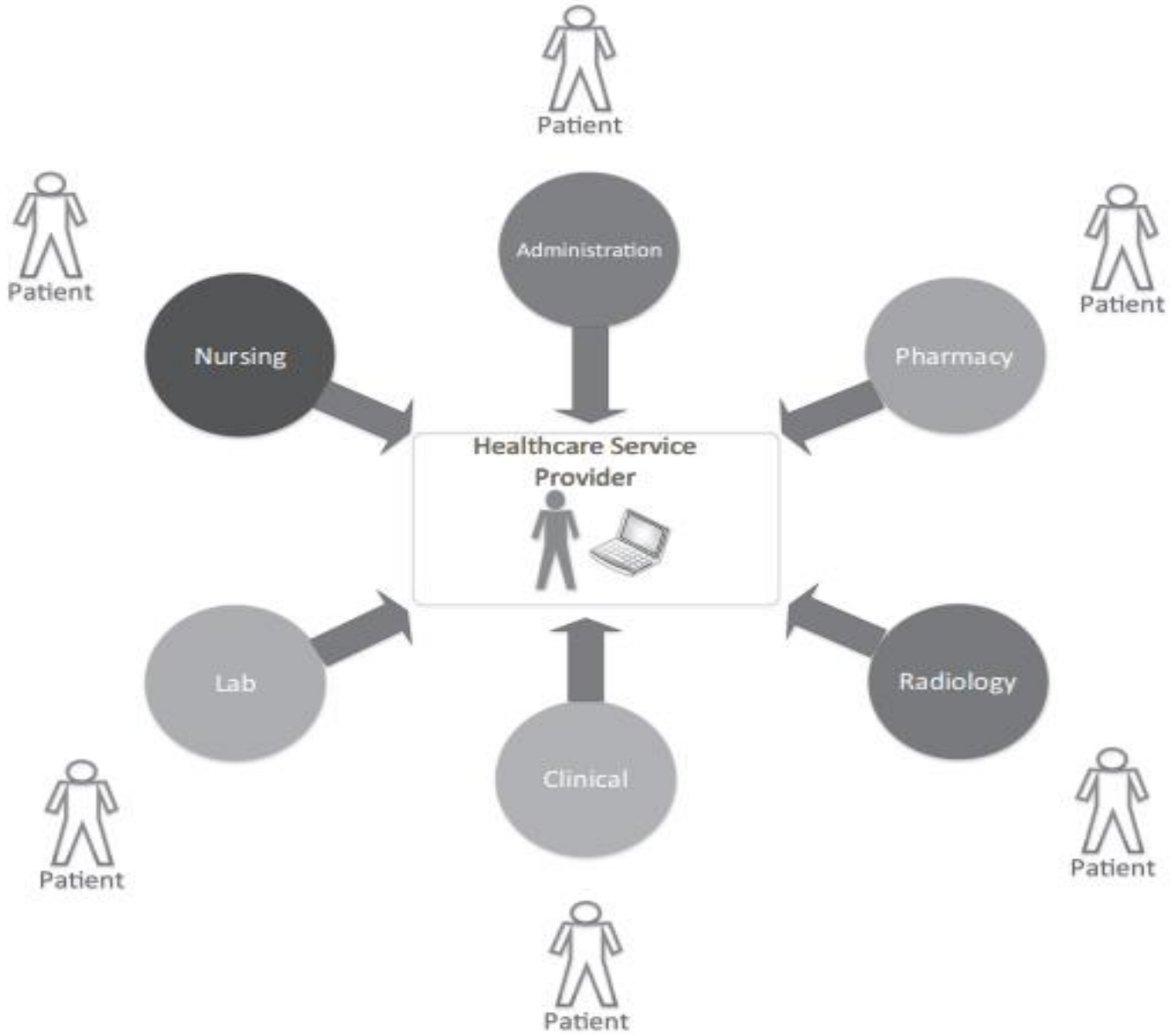



**FIGURE 2.1**: Various components of EHR.

## 1 Administrative System Components

- Administrative data are key components of the HER such as: **patient registration**, **admission**, **discharge**, and **transfer data**.

- It also includes name, **demographics, employer history, chief compliant, patient disposition, etc., along with the patient billing information**.

- **Social history data** such as marital status, home environment, daily routine, dietary patterns, sleep patterns, exercise patterns, tobacco use, alcohol use, drug use and family history data such as personal health history, hereditary diseases, father, mother and sibling(s) health status, age, and cause of death can also be a part of it.

- Apart from the **fields like "comments" or "description**," these data generally contain pairs.

- This information is used to **identify and assess a patient**, and for all other **administrative purposes.**

- During the **registration process**, a patient is generally assigned **a unique identification key** comprising of a numeric or alphanumeric sequence.
- This **key helps to link all the components** across different platforms.
- For example, **lab test data** can create an electronic record; and **another record is created from radiology results**. Both records will have the same identifier key to represent a single patient.
- Records of a previous encounter are also pulled up using this **key. It is often referred to as the medical record number or master patient index (MPI)**.
- Administrative data allows the aggregation of a person's health information for clinical analysis and research.

## 2 Laboratory System Components & Vital Signs

- Generally, laboratory systems are stand-alone systems that are interfaced to the central EHR system.

- It is a **structured data that can be expressed using standard terminology and stored in the form of a name-value pair**.

- Lab data plays an extremely important part in the clinical care process, **providing professionals the information needed for prevention, diagnosis, treatment, and health management.** About 60% to 70% of medical decisions are based on laboratory test results.

- Electronic lab data has several benefits including improved presentation and reduction of error due to manual data entry. A physician can easily compare the results from previous tests. If the options are provided, he can also analyze automatically whether data results fall within normal range or not.

- The most common coding system used to represent the laboratory test data is Logical Observation Identifiers Names and Codes (LOINC). Many hospitals use their **local dictionaries as well to encode variables**.

- Vanderbilt University Medical Center **data standardization** study found that for simple concepts such as “weight” and “height,” there were more than five internal representations.

- In different places, there are **different field names** for **the same feature and the values are stored with different units** (e.g., kilograms, grams, and pounds for weight; centimeters, meters, inches, and feet for height).

- **Vital signs** are the indicators of a **patient’s general physical condition**. It includes **pulse, respiratory rate, blood pressure, body temperature, body mass index (BMI)**, etc.

- A typical EHR system must provide the option to accommodate these kinds of variables.

## 3 Radiology System Components

- In hospital radiology departments, **radiology information systems (RIS)** are used for **managing medical imagery and associated d**ata.
- RIS is the **core database to store, manipulate, and distribute patient radiological data.**
- It uses Current Procedural Terminology (**CPT**) or International Classification of Diseases (**ICD**) coding systems to identify procedures and resources.
- Generally, an RIS consists of **patient tracking, scheduling, result reporting, and image tracking capabilities**.
- RIS is usually used **along with a picture archiving communications system (PACS),** which is a medical technology for providing economical storage and convenient access to the digital images.
- An RIS can generate **an entire patient's imagery history and statistical reports** for patients or procedures.
- Although many hospitals are using RIS, it may or may not be integrated with the central EHR system

## 4 Pharmacy System Components

- In hospitals and clinics, the **pharmacy, billing, and dispensing medications. department's** responsibility is to maintain the inventory, prescription management

- The pharmacy component in EHR will hold the **complete medication history of a patient such as drug name, dosage, route, quantity, frequency, start and stop date, prescribed by, allergic reaction to medications, source of medication**, etc.

- Pharmacists serve an important **public health role by administering immunizations** and must have the capabilities to document these services and share this information with other healthcare providers and public health organizations.

- They assure **safe and effective medication and supporting patient-centered care**.

- Pharmacies are highly automated in large hospitals. Again, it may be independent of central EHRs.

- The **Food and Drug Administration (FDA)** requires all the drugs to be registered and reported using a National Drug Code (**NDC**). Coding systems used are NDC, **SNOMED**, and RxNorm.

## 5 Computerized Physician Order Entry (CPOE)

- **Computerized Physician Order Entry (CPOE)** is a very important part of EHRs.

- It is a **system that allows a medical practitioner** to enter **medical orders and instructions for the treatment of a patient**. For example, a **doctor can electronically order services to laboratory, pharmacy, and radiology services through CPOE**.

- Then it gets propagated over a **network to the person responsible for carrying out these orders.**

- As a digital system, CPOE has **the potential to reduce medication related errors.**

- **It is possible to add intelligent rules for checking allergies, contradictions, and other alerts.**

The **primary advantages of CPOE** are the following:

- overcomes the issue of illegibility
- fewer errors associated with ordering drugs with similar names
- more easily integrated with decision support systems, easily linked to drug-drug interaction warning
- more likely to identify the prescribing physician, able to link the adverse drug event (ADE) reporting systems
- able to avoid medication errors like trailing zeros
- create data that is available for analysis, point out treatment and drug of choice,
- reduce under- and overprescribing and finally, the prescriptions can reach the pharmacy quicker.
- While ordering, a professional can view the medical history, current status report from a different module, and evidence-based clinical guidelines.

- Thus, CPOE can help in patient-centered clinical decision support.

- If used properly, CPOE decreases delay in order completion, reduces errors related to handwriting or transcriptions, allows order entry at point-of-care or off-site, provides error checking for duplicate or incorrect doses or tests, and simplifies inventory and positing of charges.

- Studies have shown that CPOE can contribute to shortened length of stay and reduction of cost.
- There are some risks involved in adopting CPOE as well. It may slow down interpersonal communication in an emergency situation.

- If each group of professionals (e.g., physicians and nurses) works alone in their workstations, it may create ambiguity about the instructions.

- These factors led an increase in mortality rate by 2.8%–6.5% in the Children's Hospital of Pittsburgh's Pediatric ICU when a CPOE system was introduced.

- Frequent alerts and warnings may also interrupt workflow. The adaptation rate of CPOE is slow. It may be partly due to physicians' doubt about the value of CPOE and clinical decision support

## 6 Clinical Documentation

- A clinical document contains the information related to the care and services provided to the patient.
- It increases the value of EHR by allowing electronic capture of clinical reports, patient assessments, and progress reports.
- A clinical document may include:

• Physician, nurse, and other clinician notes
Relevant dates and times associated with the document.
The performers of the care described
• Flow sheets (vital signs, input and output, and problems lists)
• Perioperative notes
• Discharge summaries
• Transcription document management
• Medical records abstracts
• Advance directives or living wills
• Durable powers or attorney for healthcare decisions
• Consents (procedural)
• Medical record/chart tracking
• Release of information (including authorizations)
• Staff credentialing/staff qualification and appointments documentations
• Chart deficiency tracking
• Utilization management
• The intended recipient of the information and the time the document was written.

- The sources of information contained within the document

- Clinical documents are important because documentation is critical for patient care, serves as a legal document, quality reviews, and validates the patient care provided.

- Well-documented medical records reduce the re-work of claims processing, compliance with CMS (Centers for Medicare and Medicaid Services), Tricare and other payer's regulations and guidelines, and finally impacts coding, billing, and reimbursement.

- A clinical document is intended for better communication with the providers.
-
- It helps physicians to demonstrate accountability and may ensure quality care provided to the patient.

- A clinical document needs to be patient centered, accurate, complete, concise, and timely to serve these purposes.

- The clinical document architecture (CDA) is an XML-based electronic standard developed by the Health Level 7 International (HL7) to define the structure. It can be both read by human eyes and processed by automatic software

## 2.5 Benefits of EHR

- EHRs are transformational tools.
- The scope of paper-based systems is severely limited.
- We need EHRs to **improve the quality of patient care** and **increase productivity and efficiency.**
- In terms of the overall management and costs, EHRs are a better choice.
- They also help in complying with government regulations and other legal issues.
- The benefits of EHRs are:

1. **Enhanced Revenue . . . . . .**
2. **Averted Costs . . . . . .**
3. **Additional Benefits . . . . . .**

## 1 Enhanced Revenue

- An EHR system can capture **the charges and bills for clinical services** provided, laboratory tests, and medications more accurately.
- Utilization of electronic **systems decrease billing errors**.
- They also provide a **better documentation opportunity for these services** that can be used **to resolve financial disputes**.
- **Better management of information** yield **more accurate evaluation** and increase reimbursements.
- According to experts, due to **inaccurate coding systems**, 3%–15% of a healthcare provider's total revenue is lost.
- An EHR system can be **programmed or configured to generate alerts for both patients and doctors when a healthcare service is due**. This can aid better management of collecting revenue.

- It can be used **to garner more revenues by incorporating services** like telemedicine, e-visits, virtual office visits, etc.

- It is true that all kinds of services are not possible over the Internet or telephone network, but **not all diseases will require extensive diagnosis and laboratory testing.**

- **Diseases commonly treated through telemedicine include**:

- acne, allergies,
- cold and flu,
- constipation,
- diabetes,
- fever,
- gout,
- headache,
- joint aches and pains,
- nausea and vomiting,
- pink eye,
- rashes, sinus infection, sore throat, sunburn and urinary tract infections, anxiety and depression, etc.

## 2 Averted Costs

- After adopting electronic systems, **some costs associated with the previous way of operating a business are eliminated**.

- The Center for Information Technology leadership suggested that the use of EHRs will save a total of $44 billion each year.

- Adopting EHR has the following averted costs:

- **Reduced paper and supply cost:**
  -To maintain paper-based health records an organization will require a lot of paper, printing materials, and other supplies.
  -Adopting EHR will reduce these costs.
  -After adopting EHRs, one organization estimated a reduction of 90% of paper usage within a few months.

- **Improved utilization of tests:** In electronic systems, test results are better organized. A healthcare staff no longer needs to carry the reports from one place to another. Identifying redundancy or unnecessary tests is easier. This can reduce the loss of information and ensure improved utilization of tests.

- A study by Wang et al. reports better utilization of radiology tests after adopting EHRs.

• **Reduced transcription costs**:

- An EHR **can reduce transcription costs** for manual administrative processes.

- It utilizes **structured flow sheets, clinical templates, and point-of-care documentation.**

-In a typical outpatient setting, physicians generate about **40 lines of transcription per encounter.**

- For a group of three practicing physicians, treating 12,000 patients annually at the cost of $0.11 for each transcription line results in over $50,000 per year.

- A study of fourteen solo or small-group primary care practices in twelve U.S. states reports the median transcription cost saving to be $10,800, where a minimum saving was $8,500 and a maximum was $12,000 for the year 2004–2005.

- Other related research work also describes saving $1,000–$3,000 per physician, per month.

- **Improved productivity:**

- EHR helps to improve workflows by utilizing resources more efficiently and reducing redundancies.
- As a result, the overall productivity of individuals increases.

- **Better availability of information and elimination of chart**:

- In EHR, all the charts are in digital format.
- It eliminates the need to pull, route, and re-file paper charts.
- A significant amount of effort is spent on creating, filing, searching, and transporting paper charts.
- A study estimated that the elimination of paper charts can save $5 per chart pull. It is also comparatively easier to manage digital charts.

- **Improved clinician satisfaction:**

- Electronic technology can save time by reducing the paperwork burden, which can create additional time for patient encounters and delivery of care.
- A study reports the use of EHR has reduced the physician's office visit time by 13% and a nurse's pre-exam interview time by 1 minute.
- This can improve satisfaction for professionals, which can indirectly enhance revenue.

**3. Additional Benefits**

-EHR offers **many additional benefits** that are discussed in more detail below:

- **Improved accuracy of diagnosis and care:**

-EHR provides **comprehensive and accurate patient information to physicians** that can help to quickly and systematically identify the correct problem to treat. -

-EHRs **do not just contain the patient information; they have the capability to perform computation and make suggestions.**

-They **can also present comparative results** of the standard measurements:

A U.S. national survey of doctors demonstrates the following:

- – 94% of the providers report EHR makes records readily available at the point of care.
- – 88% report that EHR produces clinical benefits for their practice.
- – 75% report that EHR allowed them to deliver better patient care.

- The gathered information can guide a physician in the emergency department to take prudent and safer actions.
- Such services are unimaginable with paper-based systems. Diagnostic errors are difficult to detect and can be fatal to a patient.

- **Improved quality and convenience of care:**

- EHRs have the **potential to improve the quality of care** by embedding options such as **Clinical Decision Support (CDS), clinical alerts, reminders**, etc.

- Research suggests that **EHRs are linked to better infection control**, **improved prescribing practices**, and **improved disease management in hospitals**.

- In such applications, **convenience is also an important measure**. EHRs greatly reduce the **need for patients to fill out similar** (or even sometimes the same) forms at each visit.

- Patients can **have their e-prescriptions ready** even before they leave the facility and can be electronically sent to a pharmacy. Physicians and staff can process claims insurance immediately.

- Following are the results of a study on the effects of e-prescribing reports:-
  -92% patients were happy with their doctor using e-prescribing.
  – 90% reported rarely or only occasionally having prescriptions not ready after going to the pharmacy.
  – 76% reported e-prescribing made obtaining medications easier. – 63%

- **Improved patient safety:**

-Just like improving the quality of care, clinical decision support systems (CDSS) and computerized physician order entry (CPOE) have the potential to improve patient safety.

-**Medication errors are common medical mistakes an**d in the United States it is responsible for the death of a person every day on average as well as injuring more than a million annually.

- Research shows that utilization of **CPOE can reduce medication errors**.
- -Medication errors can occur at any stage of the medication administration process from a physician ordering the drug, followed by the dispensing of the drug by the pharmacist, and finally the actual administration of the drug by the nurse.
- CPOE is a technology that allows physicians to act on a computerized system that introduces structure and control. Along with patient information, EHR holds the medication records for a patient.
- Whenever a new medication is prescribed, it can check for potential conflicts and allergies related to the particular medication and alert the physician.
- The system also can provide the chemical entities present in the drug and cross-reference allergies, interactions, and other possible problems related to the specific drug. Introducing technologies such as Barcode Medication Administration can make the system even more accurate.
- The Institute of Medicine (IOM) recommends CPOE and CDS as main information technology mechanisms for increasing patient safety in the future.

- **Improved patient education and participation:**

- In an EHR system, certain features can provide simplified patient education.
- EHRs can be used by the provider as a tool to illustrate procedures and explain a patient's conditions.

- It can increase a patient's participation by offering follow-up information, self-care instructions, reminders for other follow-up care, and links to necessary resources.

- Information technology affects every part of our life. In this digital era, patients may feel more comfortable with an electronic system.

**Improved coordination of care:**

- EHRs are considered essential elements of care coordination.
- The National Quality Forum defines care coordination as the following:

"Care coordination is a function that helps ensure that the patient's needs and preferences for health services and information sharing across people, functions, and sites are met over time. Coordination maximizes the value of services delivered to patients by facilitating beneficial, efficient, safe and high-quality patient experiences and improved healthcare outcomes."

- For a patient with multiple morbidities, a physician is responsible for providing primary care services and coordinating the actions of multiple subspecialists.
- According to a Gallup poll, it is a common scenario for older patients to have multiple doctors: no physician 3%, one physician 16%, two physicians 26%, three physicians 23%, four physicians 15%, five physicians 6%, and six or more physicians 11%.
- EHRs allow all clinicians to document services provided and access up-to-date information about their patient. It streamlines the transition process and knowledge sharing between different care settings. This facilitates an improved level of communication and coordination.
- Research suggests that the clinicians having 6+ months use of EHRs reported better accessing and completeness of information than clinicians without EHRs. Clinicians having EHRs have also reported to be in agreement on treatment goals with other involved clinicians.

- **Improved legal and regulatory compliance:**

- As organizations develop their systems, it is important to understand and comply with many federal, state, accreditation, and other regulatory requirements.

- A health record is the most important legal and business record for a healthcare organization. The use of an EHR system will provide more security and confidentiality of a patient's information and thus, comply with regulations like HIPAA (Health Insurance Portability and Accountability Act (HIPAA)), Consumer Credit Act, etc.

- Moreover, the Centre for Medicare and Medicaid Services (CMS) has financial incentive programs for hospitals regarding the meaningful use of health information technology.

- To receive the financial reimbursement, professionals have to meet a certain criteria and can get up to $44,000 through Medicare EHR Incentive Program and up to $63,750 through the Medicaid EHR Incentive Program.

- Adaptation of certified EHR can help providers get reimbursed

**Improved ability to conduct research and surveillance:**

- In conjunction with the direct use of EHR in primary patient care, there is an increasing recognition that secondary use of EHR data can provide significant insights.

- Using quantitative analysis of functional values, it has the potential to identify abnormalities and predict phenotypes.

- Pakhomov et al. demonstrated the use of text processing and NLP to identify heart failure patients.
- EHR data can be used to predict survival time of patients.

- Data from different EHRs can be integrated into a larger database and geo-location specific surveillance is also possible.

- **Improved aggregation of data and interoperability:**

-Standards play a crucial role in **data aggregation and interoperability** between different systems.

-EHRs **maintain standard procedure and follow defined coding system** while collecting data.

- This accommodates easier aggregation of data and greater interoperability, which offer the following benefits.

– Manage increasingly complex clinical care

– Connect multiple locations of care delivery

– Support team-based care

– Deliver evidence-based care

– Reduce errors, duplications, and delay

– Support ubiquitous care – Empower and involve citizens – Enable the move to the Personal Health Paradigm

– Underpin population health and research

– Protect patient privacy

- We need high-quality aggregated data from multiple sources in order to make evidence based decisions.

-The level of achievable interoperability using EHRs is unthinkable from paper-based systems.

-The American Medical Association recognizes that enhanced interoperability of EHRs will further help to attain the nation's goal of a high-performing healthcare system

- **Improved business relationships:**

-A healthcare provider organization equipped a with superior EHR system can be in a better bargaining position with insurers and payers compared with less equipped ones.

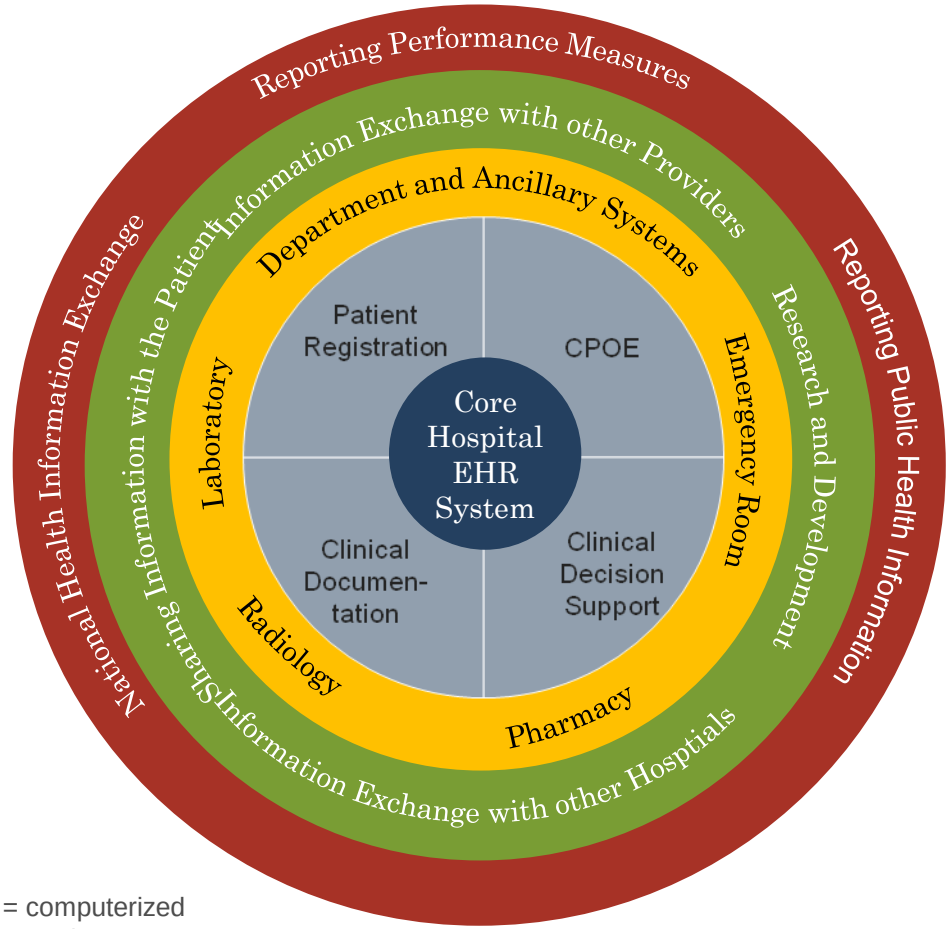
-The next generation of business professionals will expect and demand a state-of-the-art information healthcare technology system.

- **Improved reliability:**

-Data is more reliable in a digital format. Due to the reduction of storage costs, having multiple copies of data is possible.

# EHRs can facilitate communication within and outside the hospital.

**EHR Functions and Communication Capabilities**




CPOE = computerized physician order entry

**Barriers to Adopting EHR:**

- Despite of having **great potential of EHRs** in medical practice, the **adoption rate is quite slow** and faces a range of various obstacles.

- **Many other developed countries are doing far better than the United States**. Four nations (United Kingdom, the Netherlands, Australia, and New Zealand) have almost universal use (each ~90%) of EHRs among the general practitioners.

- In contrast, the United States and Canada have only around 10–30% of the ambulatory care physicians using EHRs.

- **Health informatics has been a high priority in other developed nations**, while until recently, the degree of involvement and investment by the U.S. government in EHRs has not been significant.

- Major barriers to adopting EHRs are discussed below

**1. Financial barriers:**

- Although there are studies that **demonstrate financial savings after adopting EHRs**, the reality is that the **EHR systems are expensive**.

- Several surveys report that the **monetary aspect is one of the major barriers** of adopting EHRs.

- There are mainly **two types of financial costs, start-up and ongoing**. A 2005 study suggests that the average initial cost of setting up an EHR is $44,000 (ranging from a minimum of $14,000 to a maximum of $63,000) and ongoing costs average about $8,500 per provider per year.

- **Major start-up costs include purchasing hardware and software**. In addition, a **significant amount of money is also required for system administration, control, maintenance, and support**.

- **Long-term costs** include **monitoring**, **modifying**, and **upgrading** the system as well as storage and maintenance of health records.

- Besides, **after** the **substantial amount of investment**, **physicians are worried that it could take up to several years for the return on the investment**.

**2. Physician's resistance:**

- To adopt EHRs, **physicians have to be shown that new technology** can **return financial profits, saves time, and is good for their patients' well-being.**

- Although research-based evidence is available, it is difficult to **provide concrete proof of those benefits**.

- As given in a report by Kemper et al., 58% of **physicians are without any doubt that EHR can improve patient care or clinical outcomes**.

- Finally, adopting **EHRs in a medical practice will significantly change the work processes that physicians have developed for years.**

- Besides, **physicians and staffs might have insufficient technical knowledge to deal with EHRs, which leads them to think EHR systems are overly complex**.

- Many **physicians complain about poor follow-up services** regarding **technical issues and a general lack of training and support from EHR system vendors**.

- A study reports that **two-thirds of physicians expressed inadequate technical** support as a barrier to adopting EHRs.

- Some **physicians** are also **concerned about the limitation of EHR capabilities**.
- Under **certain circumstances or as time passes**, the **system may no longer be useful**.
- Besides, **all physicians do not perform the same operations**. EHR systems have to be **customizable to best serve each purpose**.
- Surveys suggest that one of the **reasons for not adopting EHRs** is that the **physicians cannot find a system that meets their special requirements**.
- . However, an **increased effort and support from vendors may play a role in motivating physicians** towards adopting EHRs.

**3. Loss of productivity:**

- Adoption of an EHR system **is a time-consuming process**. It requires a **notable amount of time to select, purchase, and implement the system** into clinical practice.

- During this **period physicians have to work at a reduced capacity**. Also, a **significant amount of time** has to **be spent on learning the system.**

- The **improvement will depend on the quality of training, aptitude**, etc.

- The fluent **workflow will be disrupted during the transition period**, and there **will be a temporary loss of productivity**.

**4. Usability issues:**

- EHR software needs to be **user-friendly**.
- The **contents of the software must be well-organized** so that a **user can perform a necessary operation with a minimal number of mouse clicks or keyboard actions.**
- The **interface of software workflow** has to be intuitive enough.
- In terms of usability, a **comprehensive EHR system may be more complex** than expected.
- It has to **support all the functionalities in a provider's setting**.
- There might be a number of modules and submodules, so the **user might get lost and not find what he is looking for.**
- This has the **potential to hamper clinical productivity** as well as to **increase user fatigue, error rate, and user dissatisfaction**.
- **Usability and intuitiveness** in the system do not necessarily correlate to the amount of money spent.

- The Healthcare Information and Management Systems Society (HIMSS) has an **EHR usability task force**.
- A 2009 survey by the task force reported 1,237 usability problems, and the severity of 80% of them was rated “High” or “Medium”.
- Apart from the workflow usability issue, other related issues are configuration, integration, presentation, data integrity, and performance.
- The **task force defined the following principles** to follow for effective usability:
- **simplicity**, **naturalness**, **consistency**, **minimizing cognitive load**, **efficient interactions**, **forgiveness and feedback**, **effective use of language**, **effective information presentation**, and **preservation of context**.

**5. Lack of standards:**

- **Lack of uniform** and **consistent standards** hinders the EHR adoption.
- Standards play an **integral role** in enabling interoperability.
- CMS reimbursement for meaningful use **requires EHR systems** to demonstrate the ability to exchange information.
- Many of the currently used systems have utility **only for certain specific circumstances**.
- Different vendors have **developed systems in different programming languages** and database systems.
- They do not have **any defined best practice or design patterns**.
- This makes the **data exchange difficult or impossible between the systems**.

- This **lack of standardization limits the proliferation of EHRs**.
- While large **hospital systems have moved to EHRs**, many others are **skeptical** about the **available systems**.
- They **fear that the EHR software they buy now might not work** with **standards adopted** by the healthcare industry or mandated by the government later on.

## 6. Privacy and security concerns:

- Health records contain personal, diagnostics, procedures, and other healthcare related sensitive information.
- Due to the **immense importance of this information, an EHR system may be subjected to attack**.
- Some of the **medical diagnoses are considered socially stigmatized**, like sexually transmitted disease.
- **Some information** relates to direct life threats, like allergies.
- **Employers as well as insurance companies** may be interested to know more about a **patient to make unethical decisions** whether to cover a patient and/or his specific diagnosis. It can also influence **some of the hiring decisions**.
- EHRs contain information like **social security numbers, credit card numbers, telephone numbers, home addresses**, etc., which makes EHRs attractive target for attackers and hackers.
- A patient might **even be motivated to alter his or her medical records** to **get worker's compensation or to obtain access to narcotics**.

# What is privacy?

- Hard to define
- "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"
  - Alan Westin, Privacy and Freedom, 1967

- Therefore, it is important that the **privacy and security of EHRs are well maintained.**

- The most used certification for privacy and security is given by the Certification Commission for Healthcare Information Technology (CCHIT).

- The CCHIT website claims that by mid-2009, 75% of EHR products in the marketplace were certified.

- In addition to that, the Health Information Technology for Economic and Clinical Health (HITECH) Act introduced a new certification process sponsored by the Office of the National Coordination for Health Information Technology (ONC) in 2009.

- In January 2010, the **Office of the National Coordinator for Health IT (ONC)** released the interim final rule that provides an initial set of standards, implementation specifications, and certification criteria of EHR technology.

- Its requirement includes **database encryption, encryption of transmitted data, authentication, data integrity, audit logs, automatic log off, emergency access, access control**, and account of HIPPA release of information.

- Physicians **doubt the level of security** of patients' information and records. According to Simon et al., physicians are more concerned about this issue than patients.

- The **inappropriate disclosure of information might lead to legal consequences**.

- Testing the security of EHR products, a **group of researchers** showed that they were able to exploit a range of **common code-level and design-level vulnerabilities of a proprietary and an open source EHR**.

- These **common vulnerabilities could not be detected by 2011 security certification test scripts used by CCHIT**. EHRs pose new challenges and threats to the privacy and security of patient data.

- This is a **considerable barrier to EHRs proliferation**. However, this risk can be mitigated by proper technology, and maintaining certified standards with the software and hardware components

- HIPAA Privacy Rule
  - Protects individually identifiable health information
  - Mandates rules all covered entities must follow in protecting patients' privacy
  - Information that can be used to find out a person's identification, is referred to as **protected health information (PHI)**

- HIPAA Security Rule
  - Outlines safeguards to protect health information stored on a computer system or transmitted across computer networks
  - **Administrative safeguards:** the policies and procedures that protect electronic health information
  - **Physical safeguards:** the mechanisms required to protect electronic systems, equipment, and data
  - **Technical safeguards:** the automated processes used to protect data and control access to data
    - An **audit trail** is a report that shows who has accessed information and when

- HIPAA in the Age of Health Information Technology
  - A **breach** is the acquisition, access, use, or disclosure of unsecured PHI in a manner not permitted under the HIPAA Privacy Rule

**7. Legal aspects:**

- **Electronic records of medical information should be treated as private and confidential**. Various legal and ethical questions obstruct adoption and use of EHRs.

- The legal system that relies on the paper-era regulations does not offer proper guidance regarding the transition to EHRs.

- EHRs may increase the **physicians' legal responsibility and accountability**.

- With **computer-based sophisticated auditing**, it is easy to **track what individuals have done. The documentation is comprehensive and detailed in EHRs**.

- **It can both defend and expose physicians regarding malpractice**. According to a Health Affairs article, malpractice costs around $55 billion in the United States, which is 2.4% of total healthcare spending.

- A 2010 research reveals that it was unable to **determine whether the use of EHR increases or decreases malpractice liability overall**.

- HIPAA's privacy standards also present reasonable barriers to EHR adaptation

## Challenges of Using EHR Data

- The primary purpose of EHR data is to support healthcare-related functionalities.
- **As a vast amount of data is being collected every day**, the **secondary use of EHR data** is gaining increased attention in research community to **discover new knowledge**.
- The **main areas of use are clinical and transitional research**, **public health, and quality measurement and improvement**.
- Using the EHR data, we can **conduct both patient-oriented and public health research. EHR data can be used for the early detection of epidemics and spread of diseases, environmental hazards, promotes healthy behaviors, and policy development**.
- The integration of genetic data with EHRs can open even wider horizons.
- But the **data does not automatically provide us the knowledge**.
- The **quality and accuracy of the data** is an issue to be taken care of.

1. **Incompleteness:**

- Data **incompleteness or missingness** is a widespread problem while using EHR data for secondary purpose.

- **Missing data can limit the outcomes to be studied,** the number of explanatory factors to be considered, and even the size of population included.
- Incompleteness can occur due to a lack of collection or lack of documentation.

- Hersh reports the following reasons for inaccurate reporting by professionals:

– Unaware of legal requirements

– Lack of knowledge of which diseases are reportable

– Do not understand how to report

– Assumption that someone else will report

– Intentional failure for privacy reasons

- A pancreatic malignancies study using ICD-9-CM code at the Columbia University Medical Center found that 48% of the patients had corresponding diagnoses or disease documentation missing in their pathology reports.

- Authors also report a significant amount of key variables missing (see Table 2.1).

- Patients' irregularity of communicating with the health system can also produce incompleteness.

- Based on the application in hand, type of data and proportion of data that is missing, certain strategies can be followed to reduce the missingness of data.

## 2. Erroneous Data:

- EHR data **can be erroneous as well**.
- Data is collected from **different service areas, conditions, and geographic locations**.
- Data is **collected by busy practitioners and staff**.
- Therefore, the data **can be erroneous due to human errors**.
- **Faulty equipment** can also **produce erroneous data**.
- **Validation techniques should be used to both identify and correct** erroneous data.
- Both **internal and external validation measures can be applied**.
- Internal validation is a way to check the believability of the data, e.g., unrealistic blood pressure, BMI values, etc. Dates can be used to check whether the result generated before a test has taken place.
- External validation includes comparing the data with other patients or historical values.

**3. Uninterpretable Data:**

- The captured **EHR data might be uninterpretable** to a certain extent. It is closely related with data incompleteness.

- It may occur **when some part of the data is captured** but the rest is missing.

- For example, if a **specific quantitative or qualitative measurement unit is not provided with the result value**, it will be difficult to interpret.

4. **Inconsistency:**

- Data inconsistency **can heavily affect the analysis or result**.
- Data collection technologies, **coding rules, and standards** may change over time and across institutions, which may contribute to inconsistency.
- For multi-institutional studies this issue might be common, especially because different healthcare centers use different vendors for providing apparatus, softwares, and other technologies.
- A study in Massachusetts of 3.7 million patients found that 31% of patients have visited two or more hospitals in the course of five years.

**5. Unstructured Text:**

- In spite of having many defined structures **for collecting the data, a large portion of the EHR data contain unstructured text**.
- These data are present in the form of documentation and explanation.
- It is **easy to understand them for humans**, but **in terms of automatic computational methods, detecting the right information is difficult**.
- Sophisticated **data extraction techniques** like Natural Language Processing (NLP) are being used to identify information from text notes.

**6. Selection Bias:**

- In any hospital, the **patient group will mostly be a random collection**.
- It varies depending on the nature of practice, care unit, and the geographical location of the institution.
- It will not contain the **diversity of demography**.
- This is an important challenge to overcome.
- Therefore, EHR data mining findings will not be generalizable.
- This problem must be addressed while working with the secondary use of data.

## 7. Interoperability:

- **Lack of EHR interoperability** is a major impediment towards improved healthcare, innovation, and lowering costs.
- There are various reasons behind it. EHR software from commercial vendors are proprietary and closed systems.
- Most **software were not built** to **support communication with a third party and developing new interfaces** for that purpose might be a costly undertaking.
- Absence of standard also contributes to the problem.
- Many patients are not lenient towards sharing their information.
- Besides EHR systems must comply with the HIPAA Act to ensure the security and privacy of the data.
- In a recent JAMIA (Journal of the American Medical Informatics Association) article, the authors have specified 11 specific areas that present barriers to interoperability of C-CDA documents by inspecting 91 C-CDA documents from 21 technologies. In June 2014, the office of the National Coordinator for Health Information Technology (ONC) unveiled a plan for robust healthcare information sharing and aggregation and interoperability increase by 2024.

- Its three-year agenda includes “Send, Receive, Find, and Use Health Information to Improve Health Care Quality.”

- Its six-year agenda states “Use Information to Improve Health Care Quality and Lower Cost,” and finally, its 10-year agenda proposes to achieve a “Learning Health System.”

- The mentioned building blocks for attaining the goals are the following:

– Core technical standards and functions

– Certification to support adoption and optimization of health IT products and services

– Privacy and security protections for health information

– Supportive business, clinical, cultural, and regulatory environments

– Rules of engagement and governance

## Phenotyping Algorithms

- Phenotyping algorithms are combinations of multiple types of data and their logical relations to accurately identify cases (disease samples) and controls (non-disease samples) from EHR as illustrated.
- .
- Based on the structure, EHR data can be broadly divided into two parts, structured and unstructured data.

- Structured data exists in a name–value pair while unstructured data contains narrative and semi-narrative texts regarding descriptions, explanation, comments, etc.

- Structured data include billing data, lab values, vital signs, and medication information. Billing and diagnosis-related data are collected using various coding systems like ICD, CPT, and SNOMEDCT.

- These codes are important parts of the phenotyping process. ICD codes generally have high specificity but low sensitivity.

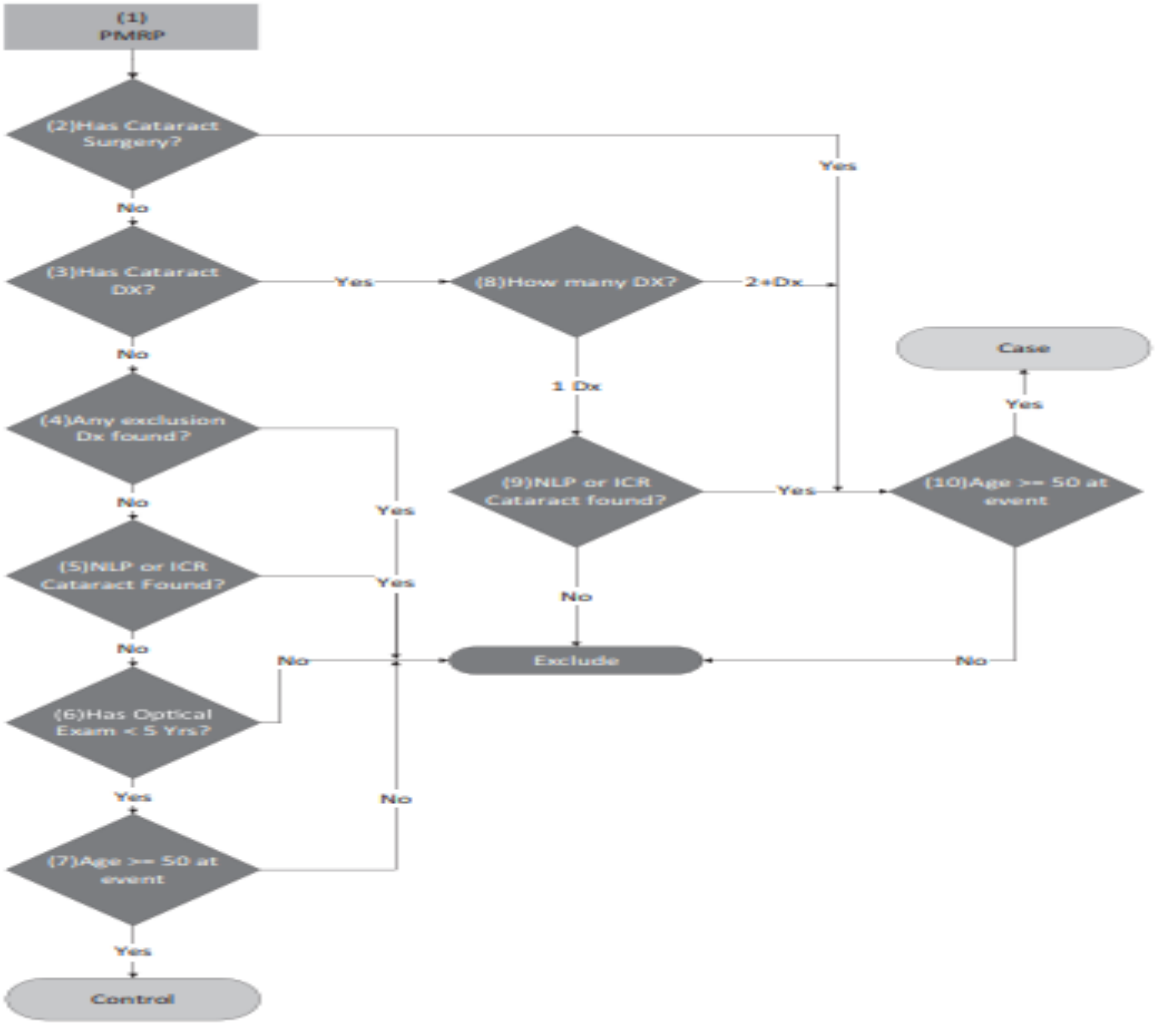



**FIGURE 2.3**: Flowchart for cataracts phenotyping algorithm taken from [98].

Table lists different characteristics of EHR data.

- The primary purpose of EHR data is to support healthcare and administrative services.
- Information is produced as a by-product of routine clinical services.
- They are not a suitable format for performing research tasks.
- They often require further processing to be used for phenotyping algorithms.
- Within existing EHR systems, querying for a particular diagnosis or lab test across all patients can be a not-trivial task.
-

- An EHR can quickly pull the information related to a patient's current medications, and easily find any test results.
- But combining different data with a temporal relationship might require manual processing of data.
- From clinical operational settings, data are often extracted and reformatted to make them more convenient and suitable for doing research, typically storing them in relational databases.
- Researchers have created a number of Enterprise Data Warehouses (EDWs) for EHR data.
- Examples include Informatics for Integrating Biology and the Bedside (i2b2), the Utah Population Database, Vanderbilt's Synthetic Derivative, etc.
- Commercial EHR vendors are also developing research repositories. For example, EPIC users can add the "Clarity" module to their system, which will convert the EHR data into SQL-based database for research purposes.

TABLE 2.2: Characteristics of Different EHR Data

| | ICD | CPT | Lab | Medication | Clinical notes |
|---|---|---|---|---|---|
| **Availability** | High | High | High | Medium | Medium |
| **Recall** | Medium | Poor | Medium | Inpatient: High Outpatient: Variable | Medium |
| **Precision** | Medium | High | High | Inpatient: High Outpatient: Variable | Medium/High |
| **Format** | Structured | Structured | Mostly | Structured | Structured |
| **Pros** | Easy to work with, good approximation of disease status | Easy to work with, high precision | High data validity | High data validity | More details about the doctors' thoughts |
| **Cons** | Disease code often used for screening, therefore disease might not be there | Missing data | Data normalization and ranges | Prescribed not necessarily taken | Difficult to process |

*Source*: Taken from Denny [106].

- To build a phenotype algorithm, first we need to select the phenotype of interest, followed by the identification of key clinical elements that define the phenotype.

- It may contain billing codes, laboratory and test results, radiology reports, medication history, and NLP-extracted information.

- The gathered information may be combined with a machine learning method.
- For example,, the authors have applied Support Vector Machine (SVM) to a both naive and well-defined collection of EHR features to identify rheumatoid arthritis cases.

- A medication record can be used to increase the accuracy of case and control identification of phenotyping algorithms.

- Patients who are believed to be controls must be having a different medication profile. They may not even have any medications prescribed to them at all.

- Sufficient dosage of a particular medication serves the confirmation that a person is having the disease of interest.

- For example, a patient treated with either oral or injectable hypoglycemic agents will be having diabetes. These medications are highly sensitive and specific for treating diabetes

**TABLE 2.3**: Phenotyping Algorithms Developed by eMERGE Network

| Phenotype | EHR data used to characterize phenotype | Institution |
|---|---|---|
| Atrial Fibrillation — Demonstration Project | CPT Codes, ICD 9 Codes, Natural Language Processing | Vanderbilt University |
| Cardiac Conduction(QRS) | CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Vanderbilt University |
| Cataracts | CPT Codes, ICD 9 Codes, Medications, Natural Language Processing | Marshfield Clinic Research Foundation |
| Clopidogrel Poor Metabolizers | CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Denny's Group at Vanderbilt, VESPA — Vanderbilt Electronic Systems for Pharmacogenomic Assessment |
| Crohn's Disease — Demonstration Project | ICD 9 Codes, Medications, Natural Language Processing | Vanderbilt University |
| Dementia | ICD 9 Codes, Medications | Group Health Cooperative |
| Diabetic Retinopathy | CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Marshfield Clinic Research Foundation |
| Drug Induced Liver Injury | ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Columbia University |
| Height | ICD 9 Codes, Laboratories, Medications | Northwestern University |
| High-Density Lipoproteins (HDL) | ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Marshfield Clinic Research Foundation |
| Hypothyroidism | CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Vanderbilt University, Group Health Cooperative, Northwestern University |
| Lipids | ICD 9 Codes, Laboratories, Medications | Northwestern University |
| Multiple Sclerosis — Demonstration Project | ICD 9 Codes, Medications, Natural Language Processing | Vanderbilt University |
| Peripheral Arterial Disease | CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Mayo Clinic |
| Red Blood Cell Indices | CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Mayo Clinic |
| Rheumatoid Arthritis — Demonstration Project | ICD 9 Codes, Medications, Natural Language Processing | Vanderbilt University |
| Severe Early Childhood Obesity | ICD 9 Codes, Medications, Natural Language Processing, Vital Signs | Cincinnati Children's Hospital Medical Center |
| Type 2 Diabetes — Demonstration Project | ICD 9 Codes, Laboratories, Medications, Natural Language Processing | Vanderbilt University |
| Type 2 Diabetes Mellitus | ICD 9 Codes, Laboratories, Medications | Northwestern University |
| Warfarin dose/response | Laboratories, Natural Language Processing | Vanderbilt University |
| White Blood Cell Indices | CPT Codes, ICD 9 Codes, Laboratories, Medications | Group Health Cooperative |

Source: Taken from [110].