Rajiur Rahman and Chandan K. Reddy, "Electronic Health Records: a Survey" in Healthcare Data Analytics, Chandan K. Reddy and Charu C. Aggarwal (eds.), CRC Press, 2014.

Contents

Chapter 1

Electronic Health Records: A Survey

1.1 Introduction

An Electronic Health Record (EHR) is a digital version of a patient's medical history. It is a longitudinal record of patient health information generated by one or several encounters in any healthcare providing setting. The term is often used interchangeably with EMR (Electronic Medical Record) and CPR (Computer-based Patient Record). 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, immunizations, laboratory data, radiology reports, personal statistics, progress notes, and billing data. The EHR system automates the data management process of complex clinical environments and has the potential to streamline the clinician's workflow. 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. An EHR system integrates data for different purposes. It enables the administrator to utilize the data for billing purposes, the physician to analyze patient diagnostics information and treatment effectiveness, the nurse to report adverse conditions and the researcher to discover new knowledge.

EHR has several advantages over paper based systems. Storage and retrieval of data is obviously more efficient in EHRs. It helps to improve quality and convenience of patient care, increase patient participation in their case, improve accuracy of diagnoses and health outcomes, and improve care coordination. It also reduces cost by eliminating the need for paper and other storage media. It provides the opportunity for research in different disciplines. In 2011, 54% of physicians had adopted an EHR system, and about three-quarters of adopters reported that using an EHR system resulted in enhanced patient care [1].

Usually, EHR is generated and maintained within an institution, such as a hospital, clinic or physician's office. An institution will contain the longitudinal records of a particular patient which has been collected at their end. The institution will not contain the records of all care provided to the patient in other venues. Information regarding to the general population may be kept in a nationwide or regional health information system. Depending on the goal, service, venue and role of the user, EHR can have different data formats, presentations and level of detail.

The remainder of this chapter is organized as follow. In section 2, a brief history of EHR development is discussed. In section 3, we list the components of EHRs. Section 4 presents a comprehensive review of existing coding systems in EHR. In section 5, we talk about EHR ontologies. In section 6, we explain the benefits of using EHRs. In section 7, we examine the barriers leading to the widespread adoption of EHRs. In section 8, we briefly explain some challenges of using EHR data. Finally, we conclude in section 9.

1.2 History of EHR

The first known medical record can be traced back to the fifth century B.C. when Hippocrates prescribed two goals for medical records [2]

- A medical record should accurately reflect the course of disease.
- A medical record should indicate the probable cause of disease.

Although these two goals are still appropriate, EHR has a lot more to offer. Modern EHR can provide additional functionalities that could not be performed using paper-based systems.

Modern day EHR first began to appear in the 1960s. Early EHRs were developed due to physicians' concerns about the increasing complexity and size of medical data. Data retrieval was much faster using digital format. Summerfield and Empey reported that at least 73 projects regarding hospitals and clinical information and 28 projects regarding storage and retrieval of clinically-relevant data were completed by 1965 [3]. In 1967, Latter Day Saints Hospitals in Utah started using Health Evaluation through Logical Programming (HELP) software. HELP is notable for its pioneering logical decision support features. In 1969, Harvard Medical School developed its own software Computer Stored Ambulatory Record (COASTER) and Duke University began to develop The Medical Record (TMR).

In 1970, Lockheed unveiled Technicon Medical Information Management System (TDS). It was implemented at El Camion Hospital in California. It came with a groundbreaking Computer Provided Order Entry (COPE) system. In 1979, Judith Faulkner, a computer programmer established Human Services Computing Inc., which developed the Chronicles data repository. The company later became Epic Systems. Epic Systems Corporation: A brief History, describes Chronicles as "Based on the single longitudinal patient record and designed to handle enterprise-wide data from inpatient, ambulatory, and payer environments."

In 1985, The Department of Veterans Affairs launched the automated data processing system Decentralized Hospital Computer Program (DHCP), which includes extensive clinical and administrative capabilities within its medical facilities. It received the Smithsonian Award for best use of Information Technology in Medicine in 1995. The current variant of DHCP is VistA (Veterans Health Information Systems and Technology Architecture). By providing care to over 8 million veterans operating in 163 hospitals, 800 clinics, and 135 nursing homes, VistA manages the largest medical system in the United States [4]. In 1983, Epic Systems launched a patient scheduling software program Cadence. This application helped clients to improve resource utilization and manage patient access. In 1988, Science Application International Corporation (SAIC) secured a \$1.02 billion dollar contract from the U.S. Government to develop a composite health care system. In 1992, Epic systems introduced the first Windows-based EHR software named EpicCare. Allscripts released the first software with an electronic prescribing solution for physicians in 1998.

From 2000 and beyond, EHR software has been increasingly trying to incorporate other functionalities to become an interactive companion for physicians and professionals. In January 2004, President Bush launched an initiative for the widespread adaptation of EHRs within the next 10 years. He said in his State of the Union Address "By computerizing health records, we can avoid dangerous medical mistakes, reduce costs, and improve care" [5]. In January 2009, in a speech at George Mason University, President Obama said "[EHRs] will cut waste, eliminate red tape, and reduce the need to repeat expensive medical tests. It just won't save billions of dollars and thousands of jobs – it will save lives by reducing the deadly but preventable medical errors that pervade our health care system" [6]. The data from a National Ambulatory Medical Care Survey (NAMCS) Physicians Workflow mail survey shows that in the year 2011, 54% physicians had adopted an HER system. About three-quarters of the adopters reported that their system meets the federal “meaningful use” criteria. Almost half (47%) of the physicians said they were somewhat satisfied, and 38% reported being very satisfied with their system. About three-quarters of the adopters reported that EHR has resulted in enhanced patient care. Nearly one-half of physicians without an EHR system at the time of the survey said they had plans for purchasing one within the next year [1].

1.3 Components of EHR

The main purpose of EHR is to support clinical care and billing. This increasingly 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 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 1.1 shows different components of an EHR system.

1.3.1 Administrative System Components

Administrative data like patient registration, admission, discharge and transfer data are key components of EHRs. It also includes name, demographics, employer history, chief compliant, patient disposition, etc, as well as 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 like 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 exists a¡name, value¿ pair. This information is used to identify and asses a patient, and all other administrative purposes. During the registration process, a patient is generally assigned a unique identification key comprising a numeric or alphanumeric sequence. This key helps to link all components across different platforms. For example, a lab test data can create an electronic record; another record is created from radiology results. Both records will have the same identifier key to represent a person. 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.

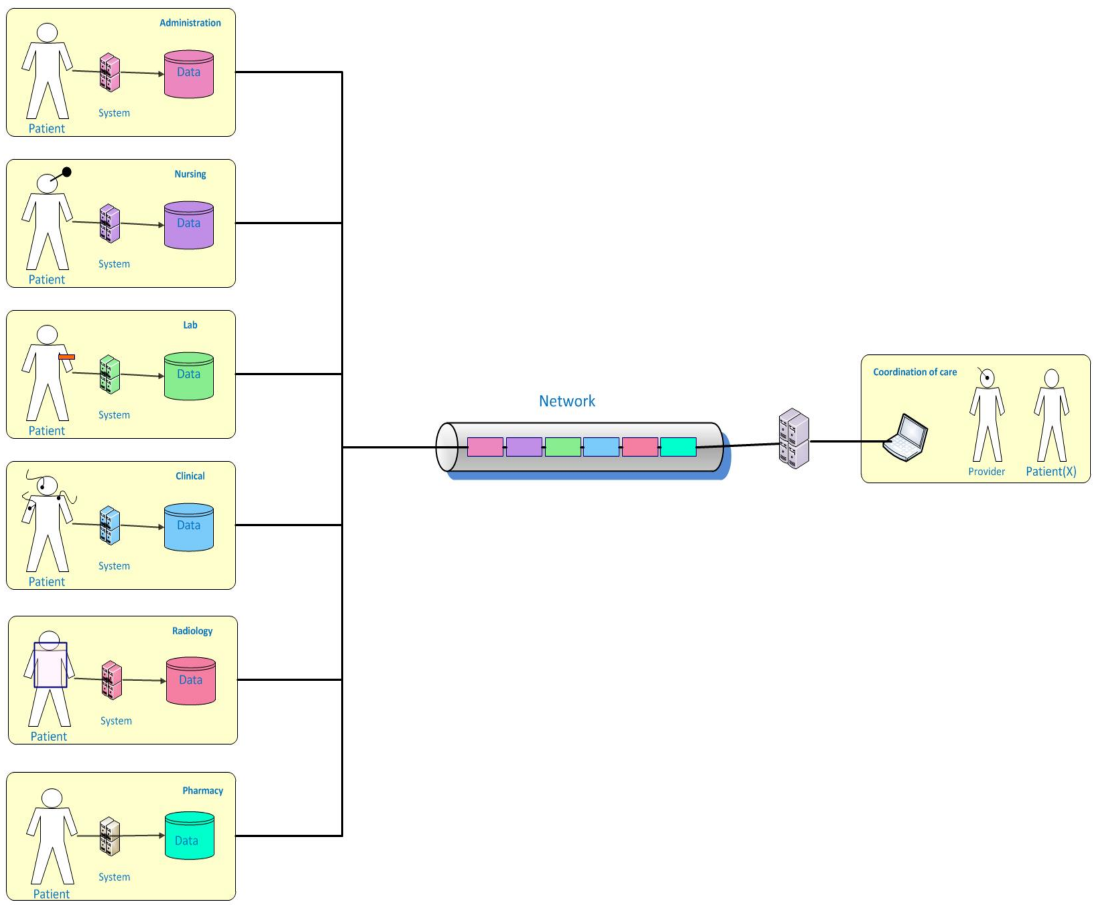

FIGURE 1.1: Components of EHR

1.3.2 Laboratory System Components & Vital Signs

Generally, laboratory systems are stand-alone systems that are interfaced to the central EHR system. It is 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 [7]. Electronic lab data has benefits including improved presentation and reduction of error due to manual data entry. A physician can compare the results from previous tests. If options are provided, he can also analyze automatically whether data results fall within normal range.

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. A 2009-2010 Vanderbilt University Medical Center data standardization study found that weight" and "height" concepts each had more than five internal representations. In different places there are different field names for the same feature and values are stored with different units (e.g. kilograms, grams, and pounds for weight; centimetres, 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 have the option to accommodate these kinds of variables.

1.3.3 Radiology System Components

In hospitals radiology departments, radiology information systems (RIS) are used for managing medical imagery and associated data. 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 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.

1.3.4 Pharmacy System Components

In hospitals and clinics, the pharmacy department's responsibility is to maintain the inventory, prescription management, billing, and dispensing medications. 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 health care 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. Food and Drug Administration (FDA) require all the drugs to be registered and reported using a National Drug Code (NDC). Coding systems used are NDC, SNOMED, and RxNorm.

1.3.5 Computerized Physician Order Entry

Computerized Physician Order Entry (CPOE) is a very important part of EHRs. It's 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 following advantages of CPOE are 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, can point out treatment and drug of choice, can reduce under and over-prescribing, and prescription reach the pharmacy quicker.

While ordering, a professional can view the medical history, current status report from different module and evidence-based clinical guidelines. Thus CPOE can help in patient-centered clinical decision support.

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 COPE can contribute to shortened length of stay and reduction of cost [8]. There are some risks involved in adopting COPE 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 instructions. These factors let an increase in mortality rate by 2.8%- 6.5% in the Children's Hospital of Pittsburgh's Paediatric ICU when a CPOE system was introduced [8].Frequent alerts and warnings may also interrupt work flow. The adaptation rate of CPOE is slow. It may be partly due to physicians doubt about the value of CPOE and clinical decision support.

1.3.6 Clinical Documentation

A clinical document contains the information related to the care and services provided to patient. It increases the value of EHR by allowing electronic capture of clinical reports, patient assessments and progress reports. A clinical document may include [9]

- 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)
- Peri-operative 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, and
- 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, 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 documents needs to be patient centered, accurate, complete, concise, and timely to serve those purposes.

The clinical document architecture (CDA) is a 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.

1.4 Coding Systems

Standards play an important role in enhancing the interoperability of health information systems and the purposeful use of EHR systems. Collecting and storing information following standard coding systems provide for better and accurate analysis of the data, seamless exchange of information, improved workflow, and reduced ambiguity. A complete health care system is complex and requires various EHR products. Different vendors have implemented standards in their own way. This result in a significant variation in the coding practices and implemented methods for which systems cannot interoperate. To create an interoperable EHR, standards are critical in four major areas

- • Applications interaction with the users
- • System communication with each other
- • Information processing and management
- • Consumer device integration with other systems and application.

Interoperability between the different EHR systems is a crucial requirement in "meaningful use of certified EHR technology" to receive incentives. That is why conforming to a standard coding system is very important. In a practical EHR, we need standards for

- Clinical vocabularies
- Healthcare message exchanges
- EHR ontologies

There are three organizations mainly responsible for developing related standards: Health Level Seven (HL7), ComiteEuropeen de Normalisation – Technical Committee (CEN TC), and the American Society of Testing and Materials (ASTM). HL7 develops healthcare related standards which are widely used in North America. CEN TC is a prominent standard developing organization works in 19 member states in Europe. Both HL7 and CEN TC collaborate with ASTM. Along with the standards developed by these organizations, EHR systems must abide by the Health Insurance Portability and Accountability (HIPPA) Act to conserve the security and privacy of patient information.

1.4.1 International Classification of Diseases

ICD stands for International Classification of Diseases which is the United Nations-sponsored World Health Organization's (WHO) official coding standard for diseases, diagnoses, health management and clinical purposes [10]. It first appeared as the International List of Causes of Death in 1893, adopted by International Statistical Institute. Since then it has been revised according to advancements in medical science and healthcare. Since the creation of WHO in 1948, it has maintained ICD. WHO published ICD-6 in 1949 and it is the first coding system where morbidity was incorporated [11]. It also included mental disorders for the first time. The U.S. Public Health Services issued International Classification of Diseases, Adapted for Indexing of Hospitals Records and Operation Classification (ICDA) in in 1959. It was revised regularly and used to classify diseases and mortality until WHO published the ninth revision of ICD.

The 1967 WHO Nomenclature Regulations specify that member nations should use the most recent ICD version for mortality and morbidity statistics. Along with storage and retrieval of epidemiological and clinical information, it allows for compilation of morbidity statistics for more than 100 WHO member nations. About 70% of the world's health expenditure in reimbursement and resource allocation is also done using ICD codes [12]. It is used to classify diseases and related problems, and provides a system of codes for a wide variety of diseases, signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or disease. It is the global foundation for providing common language in disease and health related information and statistics exchange. ICD is comprehensive and organizes information into standard groups that allows for the following [13].

- Easy storage, retrieval and analysis of health information for evidence-based decision-making.
- Sharing and comparing health information between hospitals, regions, and countries.
- Data comparison in the same location across different time periods.

1.4.1.1 ICD-9

ICD ninth revision is the most popular coding system published by WHO in 1978. It was designed to promote comparability of classification, collection, processing, and presentation of mortality statistics. Its clinical modification ICD-9-CM was published by theU.S. Public Health Services in the following year to meet statistical needs. The modified version had expanded the number of diagnostic codes and developed procedure coding system. It has more than 13,000 codes and uses more digits representing the codes than does ICD-9. It is the system that is used to encode all the diagnoses for healthcare services in United States. It is maintained by the National Center for Health Statistics (NCHS) and the Center for Medicare and Medicaid Services (CMS). Both the departments are part of the federal Department of Health and Human Service. ICD-9-CM code set is organized in three volumes and consists of tabular lists and alphabetical indices.

- Volume 1: Disease and Injuries Tabular List
- Volume 2: Disease and Injuries Alphabetical Index
- Volume 3: Procedures Tabular List and Alphabetic Index

ICD-9-CM is updated every year to keep up-to-date with medical trends and diseases. NCHS has the responsibility to update Volume 1 and 2, and CMS maintains Volume 3. Concerned parties from both the public and private sectors can propose changes to it. The major updates take effect on October 1 every year and minor updates on April 1. It is a statistical tool that converts the diagnoses and procedures into number codes. Its primary applications are

- Reporting and Research
- Monitoring the quality of patient care
- Communication and transactions
- Reimbursement, and
- Administrative uses

1.4.1.2 ICD-10

The tenth version was endorsed by WHO in 1990 during the 43rd World Health Assembly. The first full version of ICD-10 was released in 1994. The first step of implementing ICD-10 was taken by NCHS awarding a contract to the Center for Health Policy Studies (CHPS) to evaluate ICD-10 for morbidity purposes within the United States. A prototype of clinically modified ICD-10 was developed after a thorough evaluation of ICD-10 by a technical advisory panel. After strong recommendations, NCHS proceeded with implementing revised version of ICD-10-CM. During 1995-1996, further work on the enhancement of ICD-10-CM was done incorporating experiences from ICD-9-CM and through collaborating with many speciality groups like American Association of Dermatology, American Academy of Neurology, American Association of Oral and Maxillofacial Surgeons, American Academy of Orthopedic Surgeons, American Academy of Pediatrics, American College of Obstetricians and Gynecologists, American Urology Institution, and National Association of Children hospitals and Related Institutions. In 1999, ICD-10 was implemented in the United States for mortality reporting. Death statistics and data regarding leading causes of death for the years 1999 and 2000 were published using ICD-10 [**?**]. In October 2002, ICD-10 was published in 42 languages. In June/July 2003, the American Health Information Management Association (AHIMA) and American Hospital Association (AHA) jointly conducted a pilot study to test ICD-10-CM. In their study they have compared ICD-9-CM and ICD-10-CM and initial results indicated ICD-10-CM is an improvement over ICD-9-CM; and ICD-10-CM is more applicable in non-hospital environments than ICD-9-CM. Canada, Australia, Germany and others countries have their own revision of ICD-10 by adding country specific codes. The revisions are ICD-10-CA, ICD-10-AM, ICD-10-GM and so on. The standard for procedure codes ICD-10-PCS was also developed during same time frame to replace the Volume 3 of ICD-9-CM. The first revision of it was released in 1998.

ICD-9-CM is 30 years old. Many of its categories are full, and there have been changes in technology. It is also not descriptive enough. A newer coding system is needed, which would enhance reimbursement, better facilitate evaluation of medical processes and outcomes, and be flexible enough to incorporate emerging diagnoses and procedures. For example, in a scenario where a patient had a fractured left wrist and, after a month a fractured right wrist, ICD-9-CM cannot identify left versus right; additional information is required. However ICD-10-CM can report left versus right. It can characterize initial and subsequent encounter. It can also describe routine healing, delayed healing, nonunion, or malunion.

The major differences between ICD-10 and ICD-9-CM are [14]

- ICD-10 has 21 categories of diseases; in comparison ICD-9-CM has 19 categories
- ICD-10 codes are alphanumeric; in comparison ICD-9-CM codes are numeric.
- ICD-9-CM diagnoses codes are 3-5 digits in length, while ICD-10-CM codes are 3-7 characters in length.
- Total diagnoses codes in ICD-9-CM is over 14000, ICD-10-CM has 68000.
- ICD-10-PCS procedure codes are 7 characters in length. ICD-9-CM procedure codes are 3-4 numbers in length.
- ICD-10-PCS total number of codes is approximately 87000. The number of procedure codes in ICD-9-CM is approximately 4400.

The Center for Medicare and Medicaid Services (CMS) guidelines mandated a conversion from ICD-9-CM to ICD-10-CM by October 1, 2014 in the United States. Adopting a new coding system will have the following benefits

- Improve patient care. The increased detail in coding system will improve the measurement of quality, safety and efficacy of care, which will ultimately lead to improved patient care.
- Determine Severity of illness and prove medical necessity. ICD-10 codes are more granular and provide option to input the level of sickness along with complexity of disease of a patient in code based system.
- Improve research. The better and more accurate organization of code will be able to more precisely classify diseases and injuries, and correlate them with the cause, treatment, and outcome. The collected data will be less ambiguous and better defined structure of the information will make data mining easier. Information processing will be easier with newer coding system and it will open new opportunities for developing intelligent prediction system. It will also allow U.S. to conduct comparative research with other countries that are already using ICD-10.
- Lend insight to the setting of health policy. With improved data analytics afforded by ICD-10, policy makers will be able to make informed policy decisions.
- Facilitate improved public health reporting and tracking. The comprehensive coding structure will allow concerned agencies to track public health risks and trends in greater detail.

- Improve clinical, financial and administrative performance and resource allocation. The quality data can reveal essential insights. It will allow the administrators to track time and work force spent for procedures. This will help administrator to allocate resource more efficiently and achieve positive financial and managerial outcomes.

- Increase the accuracy of payment and reduce risk that claims will be rejected for incorrect coding. Reduced number of claim denials is expected due to higher specificity of within ICD-10. It will also create better electronic record of evidence to receive proper payment from government payers, insurers, hospitals, health systems and others.

- Make room for new procedures and techniques. The adaptation ability of ICD-9-CM is limited, where all the codes are consumed and has no more room for new codes. The expanded coding of ICD-10 will be able to accommodate new procedures.

- It will have other facilities like reduce hassle of audits, help preventing and detecting healthcare fraud and abuse.

1.4.1.3 ICD-11

The World Health Organization is currently working on the eleventh revision of ICD. The final publication of ICD-11 is expected by 2017 [15]. The beta draft [16] was made public online for initial comments and feedback in May 2012. This development of ICD-11 revisions is taking place in a web-based platform called iCAT, where all the concerned parties collaborate. For interested groups or people, there are options to give structured input and field testing of revised editions. It will be available in multiple languages and free to download for personal use. In ICD-11, disease entries will have definitions and descriptions of the entry and category in human readable forms. The current version ICD-10 has only title headings. There are 2400 codes in ICD-11 that are different in ICD-10 code set where, 1100 codes are related to external causes and injury [17].

Although the beta version does not support any social network platforms, it is planned to provide support social media sites such as Wikipedia, Facebook, Social Reader, Linked-in etc. The structure of definitions and other contents related to diseases and procedures will be defined more accurately. It will be more compatible with EHRs and other technologies. .

1.4.2 Current Procedural Terminology

CPT – Current Procedural Terminology— is a set of medical codes developed, maintained and copyrighted by the American Medical Association (AMA). CPT codes are a list of descriptive terms, guidelines, and identifying codes of medical, surgical, and diagnostic services designed to provide uniform communication language among physicians, coders, patients, accreditation organizations, and payers for administrative, financial, and analytic purposes.

It was first created by the AMA in 1966. The first edition contained mainly surgical codes. A significant development took place for second the edition which was published in 1970. The second edition contained 5 digits instead of 4 digits, and it included lab procedures. In 1983, the Health Claim Financial Administration (HCFA) which is now known as the Center for Medicine and Medicaid Services (CMS) ,merged its own Common Procedure Coding System (HCPCS) with CPT and mandated CPT would be used for all Medicare billing. Every year the new version is released in October. The Healthcare Common Procedures Coding System (HCPCS, often pronounced as "hick picks") is another set of codes developed by AMA based on CPT. Although the CPT coding system is similar to ICD-9 and ICD-10 it describes the treatment and diagnostic services provided while ICD codes describe the condition or the disease being treated. CPT is used only in inpatient settings.

1.4.3 SNOMED-CT

Systematized Nomenclature of Medicine – Clinical Terms is a comprehensive, computer-processible, multilingual clinical and healthcare terminology, originally created by the College of American Pathologists (CAP). SNOMED was started as Systematic Nomenclature of Pathology (SNOP) in 1965 [18]. It was enhanced further and SNOMED was created in 1974. It had two major revisions in 1979 and 1993. In 1999 SNOMED-CT was created by the merger of SNOMED Reference Terminology (SNOMED-RT) developed by the CAP and Clinical Terms Version 3 (CTV3) developed by the National Health Services of the United Kingdom. The merged version first released in 2002. SNOMED-RT had a vast coverage of medical specialities with over 12000 concepts. It was designed for the retrieval and aggregation of healthcare information produced by multiple organizations or professionals. The strong suite of CTV3 was its coverage of terminologies for general practice. With more than 200000 concepts, it was used to store primary care encounter information and patient-based records [19]. Currently SNOMED has more than 311,000 concepts with logic-based definitions organized into a hierarchy. In July 2003 the National Library of Medicine (NLM) on behalf of the U.S. Department of Health and Human Services signed a contract with CAP to make SNOMED-CT available for users. Since April 2007, it has been owned, maintained and distributed by a newly formed Denmark based non-profit organization, named International Health Terminology Standards Development Organization (IHTSDO) [9]. CAP collaborates with IHTSDO and continues to provide support SNOMED-CT operations. More than 50 countries use SNOMED-CT.

SNOMED-CT is a valuable part of EHR. Its main purpose is to encode medical and healthcare related concepts and support recording of data. It provides a consistent way to store, index, retrieve, and aggregate clinical data across different sites. It also helps to organize data in a more meaningful way and reduce the variability of the data collection and management process. Its extensive coverage includes clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other etiologies, substances, pharmaceuticals, devices, and specimen [20].

SNOMED-CT has a logical and semantic relationship between concepts. It has a multi-axial hierarchy which allows different level of details in the information. Its extensible design permits the integration of national, local and vendor specific requirements. It primarily consists of four components

- Concept Codes: numerical codes to identify terms
- Descriptions: textual descriptions of the concept codes
- Relationships: represents related concept codes
- Reference Sets: used for grouping concepts codes or descriptions. Support cross mapping to other classification standards.

SNOMED-CT can be mapped to other well-known terminologies like ICD-9-CM, ICD-10, and LOINC. Renowned standards like ANSI, DICOM, HL7, and ISO are supported by it. In a joint project with WHO it is providing insights for the upcoming ICD-11.

SNOMED-CT has some basic differences from ICD. It is mainly a terminology system where ICD is a classification system. SNOMED-CT is designed to encode and represent data for clinical purposes [21]. Information coded with ICD is used for statistical analysis, epidemiology, reimbursement, and resource allocation. SNOMED-CT facilitates the information input into the EHR and provides standardization for primary data purposes while ICD codes enable retrieval for secondary data purposes.

1.4.4 LOINC

Logical Observation Identifiers Names and Codes (LOINC) is a universal code system for identifying laboratory observations and clinical test results. In response to the demand for electronic clinical data, it was created in 1994 by Regenstrief Institute Inc., an Indianapolis-based non-profit research organization affiliated with Indiana University. It was originally called Laboratory Observations, Identifiers, names and Codes and development was sponsored by NLM and other government and private agencies. Original sources of information include [22] the following:

- Silver book for International Union of Pure and Applied Chemistry
- International Federation of Clinical Chemistry
- Textbooks of pathology
- EuCliD (European Clinical Database)
- Expertise and work of the LOINC members

LIONIC coding system helps to improve the communication of information. In January 2009, Regenstrief Institute released a Windows operating system based mapping software called Regenstrief LOINC Mapping Assistant (RELMA) where codes can be searched and local codes can be mapped to LOINC database. The current version of LOINC is LOINC 2.46 released in December 2013. With more than 600 new users per month, it has 27000 users in 158 different countries. LOINC vocabulary continues to grow till today.

Each LOINC record represents a single test result. A record consists of six fields [23]

- Component: what is measured of evaluated (example: glucose, hemoglobin)
- Kind of property: characteristics of the component that is measured (example: mass, length, concentration, volume, time stamp etc.)
- Time: observation period of the measurement.
- System: the specimen or the substance, in context of which the measurement was done (example: blood, urine)
- Scale: the measurement scale. (example: quantitative, nominal, ordinal, or narrative)
- Method (optional): the procedure which was performed for measurement.

Certain parameters and descriptors related to the test are explicitly excluded in LOINC from observation name. They are made as fields of test/observation report message [22]. These fields are

- The instrument used for testing
- Fine details of the sample or the site of collection
- The priority of the test
- Who verified the result

- Size of the sample
- Place of testing

LONIC's overall organization is divided into four categories, laboratory, clinical, attachments, and surveys. The laboratory component is further divided into sub-categories such as chemistry, hematology, serology, microbiology (includes parasitology and virology), and toxicology. The clinical attributes are vital signs, hemodynamics, intake/output, EKG, obstetric ultrasound, cardiac echo, urologic imaging, gastroendoscopic procedures, pulmonary ventilator management, and other clinical observations [22]. It also contains information about nursing diagnoses and, nursing interventions.

1.4.5 RxNorm

RxNorm is a drug vocabulary maintained and distributed by the National Library of Medicine [24]. It assigns standard names to the clinical drugs and drug delivery devices available in the United States. It is used as a basis for the capture and presentation of drug related information in EHRs. In 2001 NLM started to develop RxNorm for modeling clinical drugs in the Unified Medical Language System in consultation with HL7 vocabulary technical committee and the Veterans Administration [25]. It was developed to standardize the medication terminology that would reduce the missed synonymy in clinical drugs [26]. Additional goal was it would facilitate electronic capture of related data, improve interoperability by supporting information exchange across platforms and systems, development clinical decision support and provide opportunity for research.

RxNorm follows a standard for naming drugs. The normalized name of a drug include following components [25]

- IN – Ingredient of the drug.
- DF – Dose form of the drug
- SCDC – Semantic clinical drug component. It represents the ingredients and strength.
- SCDF – Semantic clinical drug form. It represents the ingredient and dose form.
- SCD – Semantic clinical drug. It represents the ingredient, strength and dose form
- BN – Brand name. This is the formal name for a group of drugs containing a specific active ingredient.
- SDBC – Semantic branded drug component. It represents branded ingredient and strength.
- SBDF – Semantic branded drug form. It represents the branded ingredient and dose form
- SDB – Semantic branded drug. It represents the branded ingredient, strength and dose from.

RxNorm organizes drugs by concept. A concept is a set of names with similar meaning at a specific level of abstraction. It can distinguish similar drugs from different providers using concepts. The concepts and relationships between each other form a semantic network.

1.4.6 ICF

The International Classification of Functioning, Disability and Health, commonly known as ICF, is a classification of health related components of function and disability. ICF concentrates on the functionality and body structure of people with a given health condition or disability rather than diagnosis or diseases. It does not account for the cause of disability. It is a unified and standard framework first developed by the World Health Organization (WHO) in 1980 [27]; initially it was known as International Classification of Impairments, Disabilities, and Handicaps (ICIDH). After years of coordinated revision, in May 2001, the 191 member states of WHO agreed to adopt ICF as the standard coding method of functioning and disability. In June 2008, the American Physical Therapy Association (APTA) joined WHO for endorsing ICF. ICF is the only method of its kind. It has been developed and tested for applicability in more than 40 countries.

Body functions and disability can be viewed as interactions between health condition and personal and environmental factors. ICF has mainly two parts, Functioning and disability, and Contextual factors. It can be categorized into further sub parts. The components of are listed below [28]

- Functioning and disability
  - Body functions
    - Mental functions
    - Sensory functions and pain
    - Voice and speech functions
    - Functions of the cardiovascular, hematological, immunological and respiratory systems
    - Genitourinary and reproductive functions
    - Neuromusculoskeletal and movement-related functions
    - Functions of the skin and related structures
  - Body structures
    - Structure of the nervous system
    - The eye, ear, and related structures
    - Structures involved in voice and speech
    - Structures related to cardiovascular, immunological and respiratory systems
    - Structures related to digestive, metabolic and endocrine systems
    - Structures related to genitourinary and reproductive systems
    - Structures related to movement
    - Skin and related structures
  - Activities and participation
    - Learning and applying knowledge
    - General tasks and demands
    - Communication
    - Self-care
    - Domestic Life
    - Interpersonal interactions and relationships
    - Major life areas
    - Community, social and civic life

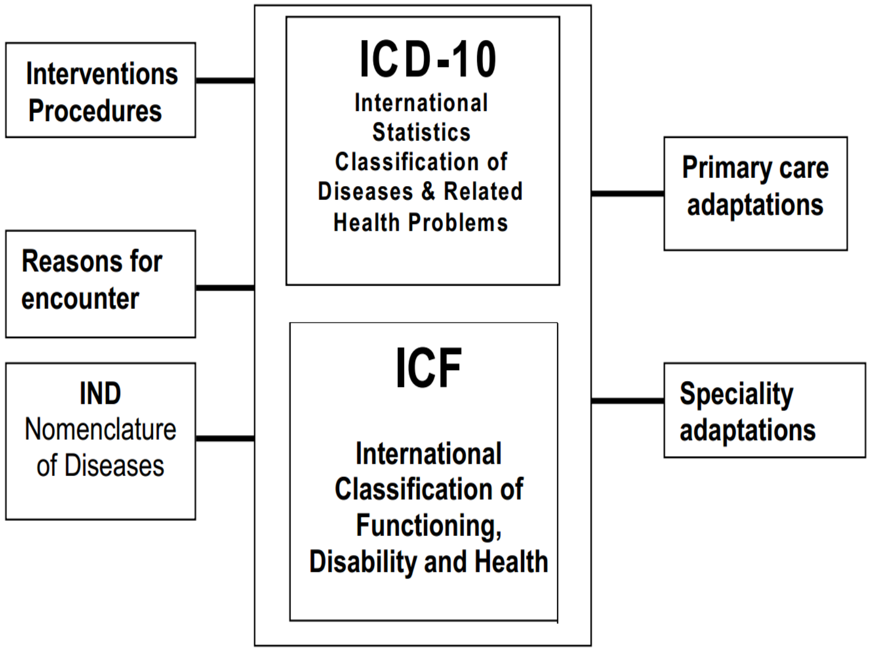

FIGURE 1.2: WHO Family of International Classifications

- Contextual factors
  - Environmental factors
    - Products of technology
    - Natural environment and human-made changes to environment
    - Support and relationships
    - Attitudes
    - Service, systems and policies
  - Personal factors
    - Gender
    - Age
    - Coping styles
    - Social background
    - Education
    - Profession
    - Past and current experience
    - Overall behavior pattern
    - Character and other factors

ICF complements WHO's classification of disease scheme, ICD-10. ICD contains diagnosis and health condition related information, but not functional status. Together they constitute the WHO Family of International Classifications (WHO-FIC) shown in figure 1.2 [29].

Diagnosis is used to define cause and prognosis of diseases, but alone it does not predict service needs, length of hospitalization, level of care of functional outcomes. Nor it can accurately provide support for disability. ICF allows incorporating all aspects of a person's life. The current ICF creates a more understandable and comprehensive profile of health forming of a person instead of focusing health condition [30]. It is used as a clinical, statistical, research, social policy, and educational tool. A common misconception about ICF is it deals with only the disable people. However, ICF has some limitations regarding the ability to classify the functional characteristics of developing children [31].

1.4.7 Diagnoses Related Group

Diagnosis Related Groups (DRG) are a patient classification scheme which group related patients and relate these groups with the costs incurred by the hospital. DRGs divide diagnosis and illness into 467 categories identified in ICD-9-CM [32]. The 467th group is "ungroupable". The classification is based on patient's principal diagnosis, ICD diagnoses, gender, age, sex, treatment procedure, discharge status, and the presence of complications or comorbidities.The goals of developing DRGs were to reduce healthcare cost, improve quality of care and efficiency of the hospitals. DRGs are by far the most important cost control and quality improvement tool developed [33].

It was first created at Yale University with the support from the Health Care Financing Administration, now known as the Center for Medicine and Medicaid Service (CMS). In 1980, it was first implemented in a small number of hospitals of New Jersey [34]. It is used to define the reimbursement amount of hospitals from Medicare. Medicare pays hospitals per patient and efficient hospitals receive better incentive. DRGs help to decide the efficiency of the hospital.

1.4.8 UMLS

The Unified Medical language System (UMLS) is a collection of comprehensive biomedical concepts and ontologies. It was developed by the US National Library of Medicine (NLM) in 1986. It provides development of computer based systems that can behave as through they understand the biomedical and health concepts [35]. It is intended to be mainly used by medical informatics professionals. NLM maintains and distributes UMLS knowledge sources (database) and related software tools for developers to build enhanced electronic information system that can create process, retrieve, integrate and/or aggregate health and biomedical related information. The knowledge sources of UMLS are as follows [36]

- Metathesaurus
  - Source vocabularies
  - Concepts
- Relationships, Attributes
  - Semantic Network
  - Semantic Types (categories)
  - Semantic relationships
- Lexical Resources
  - SPECIALIST Lexicon

  - Lexical tools

Metathesaurus is a very large, multi-purpose, and multi-lingual vocabulary database. It contains health and biomedical related concepts their various names and the relationships among them. It has 126 vocabularies in 17 languages [24]. It clusters similar terms into a concept. The semantic network provides consistent categorization of concepts defined in Metathesaurus. The network contains information regarding basic semantic types/categories which may be assigned to concepts and relationships between semantic types. In the network, the semantic types are nodes and the relationships are links between them. In the current version of semantic network, there are 135 semantic types and 54 relationships [35]. The SPECIALIST Lexicon provides the lexical information needed for the SPECIALIST natural language processing tool.

1.4.9 DICOM

The Digital Imaging and Communications in Medicine (DICOM) is a medical imaging standard. It determines the data exchange protocol, digital image format, and file structure for biomedical images and related information [37]. DICOM was developed by the American College of Radiology (ACR) and National Electric Manufacturers Association (NEMA). The first version ACR/NEMA 300 was released in 1985. DICOM is generally used in following application areas [37]

- Network image management
- Network image interpretation management
- Network print management
- Imaging procedure management
- Off-line storage media management.

DICOM allows the integration of scanners, servers, workstations, printers, and network hardware into a Picture Archiving and Communication Systems (PACS).It has been extensively used by the hospitals and other organizations. It provides a widely accepted foundation for medical imaging standards. It promotes interoperability between radiology systems.

1.5 Ontologies

In a modern EHR system, a patient's medical records are stored in different locations. Sharing of patient's information among medical professionals is extremely important to provide better patient care. Lack of interoperability is a major barrier to attaining this goal. Syntactic interoperability can assure the exchange of information structure but does not guarantee the meaning will be interpreted identically. Semantic interoperability can ensure the understanding of the shared information at the level of formally defined domain concepts [38]. Two systems will be semantically interoperable if and only if each can carry out tasks using data taken from other as seamlessly as using its own data. Ontologies can help to achieve semantic interoperability between EHR systems by defining clinical terminologies. In information systems, ontology represents a set of domain concepts which [39]

- reflects the properties of the objects within its domain in a way that obtains a systematic correlation between reality and representation itself,
- is intelligible to a domain expert,
- is formalized to support automatic information processing.

Ontology serves the purpose of automatic navigation of knowledge. It is a specification of a representational vocabulary for a shared domain [40]. Generally, ontologies describe following concepts

- Classes (things)in the domain of interest
- The properties (or attributes) those things may have
- The relationships that can exist among things

1.6 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 overall management and costs, EHRs are the better choice. They also help to comply with government regulations and other legal issues. The benefits of EHRs are described below

1.6.1 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 [41].They also provide better documentation opportunity for these services which can be used to resolve financial disputes.Better management coding and more accurate evaluation, increase reimbursements. According to experts, due to inaccurate coding systems, 3-15% of a healthcare provider's total revenue is lost [42].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 & flu, constipation, diabetes, fever, gout, headache, joint aches & pains, nausea & vomiting, pink eye, rashes, sinus infection, sore throat, sunburn and urinary tract infections, anxiety & depression etc.

1.6.2 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 total $44 billion each year [43]. Adopting EHR has the following averted costs [44]

- **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 [45].

- **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 [41].

- **Reduced Transcription Costs:** An EHR can reduce transcription costs for manual administrative processes [46] [47]. 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 $50000 per year [46]. A study of fourteen solo or small-group primary care practices in twelve U.S. states reports median transcription cost saving $10800, where minimum saving was $8500 and maximum $12,000 for the year 2004-05 [47]. Related research describes saving $1000-$3000 per physician, per month [48].

- **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 [46]. Significant amount effort is spent on creating, filing, searching, and transporting paper charts [49]. Study estimates that elimination of paper chart can save $5 per chart pull [41]. 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 encounter and deliver care [3]. A study reports the use of EHR reduced a physician's office visit time by 13% and a nurse's pre-exam interview time by 1 minute [50]. This can improve satisfaction for professionals, which can indirectly enhance revenue.

1.6.3 Additional Benefits

: EHR offers many additional benefits. They are discussed in the following sections

- **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 don't just contain the patient information; they have the capability to perform computation and make suggestions. They can also present standard measurements comparative results. A U.S. national survey of doctors demonstrates the following [51]

  - 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. A new study suggests that EHR can help to identify potential diagnostic errors in primary care by using certain types of queries (triggers) [52].

- **Improved Quality and Convenience of Care:**EHRs have the potential to improve the quality of care by embedding options like Clinical Decision Support (CDS), clinical alerts, reminders, etc. Research suggests that EHRs are linked to better infection control [53], improved prescribing practices [10], and improved disease management [42] in hospitals. Convenience also matters. They greatly reduce the need for patients to fill out the same and different 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 claim insurance immediately. Following are the results of a study on the effects of e-prescribing reports [54]
  - 92% patients were happy with their doctor using e-prescribing.
  - 90% reported rarely or only occasionally having prescription not ready after going to pharmacy.
  - 76% reported e-prescribing made obtaining medications easier.
  - 63% reported fewer medication errors.
- **Improved Patient Safety:** : Just like improving 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 and responsible for the death of one person every day as well injury more than a million in United States annually [55]. Research shows that utilization of CPOE can reduce medication errors [56] [57]. 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 actual administration of the drug by the nurse. CPOE is a technology that allows physicians to act on a computerized system which 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 like Barcode Medication Administration can make the system even more accurate. The Institute of Medicine (IOM) recommends COPE and CDS as main information technology mechanism for increasing patient safety in future [58].
- **Improved patient education and participation:** In an EHR system, certain features can provide simplified patient education [42]. EHRs can be used by the provider as a tool to illustrate procedures and explain patient's conditions. It can increase 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 parts of our life. Information technology is affecting 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 define care coordination as following [59]

"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 sub-specialists [60]. 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% [61]. 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 [62]. Research suggest that clinicians having 6+ months of EHRs use more frequently reported timely access the complete information than clinicians without EHRs. Clinicians having EHRs have also reported to be in agreement on treatment goals with other involved clinicians [63].

- **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, Consumer Credit Act etc. Moreover, the Center 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 certain criteria and can get up to $44000 through Medicare EHR Incentive Program and up to $63,750 through the Medicaid EHR Incentive Program [64]. 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 increasing recognition that secondary use of EHR data can provide significant insights [65]. Using quantitative analysis of functional values it has the potential to identify abnormalities and predict phenotypes. Pakhomove 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 [66]. 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 [67]

  – 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 high-performing health care system.

- **Improved business relationships:** A health care provider organization equipped 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 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.

1.7 Barriers to Adopting EHR

Despite of having great potential of EHRs in medical practices, 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 (the, United Kingdom, the Netherlands, Australia, and New Zealand) have almost universal use (each ¿90%) of EHRs among general practitioners. In contrast, the U.S. and Canada have 10-30% ambulatory care physicians using EHRs [68]. 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 for EHRs has not been significant. Major barriers to adopting EHRs are discussed below

- **Financial barriers:**Although there are studies that demonstrate financial savings after adopting EHRs, the reality is EHR systems are expensive. Several surveys report that the monetary aspect is one of the major barriers of adopting EHRs [69] [70] [71] [72] [73] [74] [75]. 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 $14000 to a maximum of $63,000) and ongoing cost average about $8,500 per provider per year [47]. 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 are encompassed monitoring, modifying and upgrading the system as well as and storage and maintainenance of health records. Besides, after the substantial amount of investment, physicians worry that it could take years for the return on the investment.

  EHRs are not the only electronic system that exists in any healthcare provider like practice management. There might be an old system which also needs integration into the new system. It is important that an EHR system is integrated into other systems, and this integration can be very expensive. Surveys show that due to the high financial investment required, HER adaptation was far higher in large physician practices and hospitals [76].

- **Physician's resistance:**To adopt adopting EHRs, physicians have to be shown that new technology makes money, saves time and is good for their patients. Although research based evidences is available, it is difficult to provide substantial proof of those benefits. As reported by Kemper et al., 58% of physicians without any EHR doubt that it can improve patient care or clinical outcomes [75]. Finally, adopting EHRs in a medical practice will significantly change the work processes which 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 [71]. A study by Simon et al. reports two-thirds of physicians expressed inadequate technical support as a barrier to adopting EHRs [74]. 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 [70] [73]. 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 physicians cannot find a system that meets their special requirements [70] [71] [72] [74] [77] [**?**]. However, increased effort and support from vendors may play a role in motivating physicians towards adopting EHRs.

- **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. 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 in the transition period, and there will be temporary loss of productivity [78].

- **Usability Issues:**RHR 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 key-board 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 sub-modules, so user might get lost and not find what he is looking for. This has the potential to hamper clinical productivity as well as-, 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 1237 usability problems, and the severity of 80% of them was rated "High" or "Medium" [79]. 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 [80]: simplicity, naturalness, consistency, minimizing cognitive load, efficient interactions, forgiveness and feedback, effective use of language, effective information presentation and preservation of context

- **Lack of standards:**Lack of uniform and consistent standards hinders 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 the currently used systems are of a local nature. Different vendors have developed systems in different programming languages and database systems. They don't have any defined best practice or design patterns. This makes the data exchange difficult or impossible between the systems [72] [73] [77]. This lack of standardization limits the proliferation of EHRs [81]. While large hospital systems have moved to EHRs, many others are skeptical about 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.

- **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 hire or cover a patient. EHRs contain information like social security number, credit card numbers, telephone numbers, home addresses etc. which makes EHRs attractive target for attackers. Even a patient might be motivated to alter his or her medical records to get worker's compensation or to obtain access to narcotics. 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 [82]. In addition to that, the 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 ONC released the interim final rule which provides initial set of standards, implementation specifications, and certification criteria of HER 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 [83]. 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 [73]. 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 [84]. These common vulnerabilities could not be detected by 2011 security certification test scripts 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, proper technology, and maintaining certified standards with the software and hardware components, this risk can be mitigated.

- **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 which relies on the paper-era regulations does not offer proper guidance regarding the transition to EHRs. EHRs may increase physicians' legal responsibility and accountability [85]. With computer based sophisticated auditing, it is easy to track what individuals have done. The documentation is comprehensive and detail in EHRs. It can both defend and expose physicians regarding malpractice. According to a Health Affair article, malpractice costs around 55 billion dollars in

U.S 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 [85]. HIPAA's privacy standards also present reasonable barriers to EHR adaptation.

1.8 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 behaviours, and policy development. The integration of genetic data with EHRs can open even wider horizons. The secondary use poses various challenges with the data.

First, data in EHRs can be incomplete [86] [87] and [88]. Missing data can occur due to a lack of collection or lack of documentation [89]. Missing data can limit the outcomes to be studied, the number of explanatory factors to be considered, and even the size of population included [86]. When a patient does not have any symptom/comorbidity, a lack of documentation is common. 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 [89].

Secondly, EHR data can be erroneous and inconsistent as well. Data is collected from different service areas, conditions, and geographic locations. Data is collected by busy practitioners and stuffs. Therefore, the data can be erroneous due to human error. Besides, data collection technologies and standards may change over time and across institutions, which can contribute to inconsistency. Not all the healthcare centers have the same vendor for providing a particular apparatus. If proper standards are not followed during development, it can create interoperability issues.

Although having many defined structures for collecting the data, a large portion of the EHR data contains unstructured text. These data is present in the form of documentation and explanation as a supplement. It is easy to understand them as a human, but in terms of 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 [90].

Finally, selection bias is an important challenge to overcome. In any hospital, the patient set 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. Therefore, EHR data mining findings will not be generalizable. This problem must be addressed while doing the secondary use of data.

1.9 Phenotyping Algorithms

EHR systems are gaining nation-wide popularity in United States recently due to "Meaningful Use" legislation and reimbursement [91]. It is being widely installed in hospitals, academic medical centers, outpatient clinics throughout the nation. Besides healthcare benefits like improved patient care, safety and reduced costs, it creates great opportunity for clinical and translational research. Instead of time-consuming and costly manual chart reviews of voluminous medical information, electronic data can be processed quickly. Integrating patient's genetic information with EHR can create a new era of genomic research. Collecting DNA data for genomic discovery is time consuming and expensive. These barriers prohibit experiments to perform on large cohorts.

The primary purpose of EHR data is to support healthcare and administrative services. Information is produced as a byproduct of routine clinical services. They are not in ideal format for research. 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 patient's current medications, and easily find any test results. But combining different with temporal relationship might require manual processing of data. From clinical operational settings, data are often extracted reformatted to make it more convenient for 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) [92], the Utah Population Database [93], Vanderbilt's Synthetic Derivative [94] etc. Commercial EHR vendors are also developing research repositories. For example EPIC users can add "Clarity" module to their system which place EHR data into SQL-based database for research purposes.

Recently, researchers have been linking EHR data with biological databanks (biobanks). The most popular biobanks are the collection of DNA samples. Hospitals and clinics can collect DNA samples from a patient's blood used in routine tests. The Personalized Medicine Research Population (PMRP) project in Marshfield Clinic has a biobanks of 20000 individuals [95]. Similar DNA biobanks exist at eMERGE Network sites, Northwestern University, Geisinger Health System, Mount Sinai School of Medicine and other places.

Based on the structure, EHR data can be broadly divided into two parts, structured and unstructured data. Structured data exists in name-value pair while unstructured data contains narrative and semi-narrative texts regarding descriptions, explanation, comments etc. Structure data include billing data, lab values, vital sings, and medication information. Phenotyping algorithms are logically related multiple types of data that are used to accurately identify cases (disease samples) and controls (non-diseased samples) from EHR. Billing and diagnosis related data are collected using various coding systems like ICD, CPT, and SNOMED-CT. These codes are important parts of phenotyping process. ICD codes generally have high specificity but low sensitivity [96]. Figure 3 lists different characteristics of EHR data [97].

Studies have shown that CPT codes can accurately predict occurrence of a given procedure [98]. The standard terminology codes for lab tests are LOINC. On the other hand, clinical notes are in free-text format. To be used for phenotyping algorithms, it has to undergo subsequent text processing. Some procedure and test result may also exist as a combination of structured and unstructured form. For example an electro-cardiogram report typically contains structured interval durations, heart rates, and overall categorization, along with a narrative text of cardiologist's interpretation of the result [99].

Before the use of a phenotyping algorithm, data has to be normalized to standard repre-

	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 normalizat ion and ranges	Prescribed not necessary taken	Difficult to process

FIGURE 1.3: Summary of Common EHR data

sentation. Natural Language Processing (NLP) based tools have gained much popularity to extract structured information from free text. Several studies have shown that coded data are not sufficient or accurate to identify disease cohorts [100] [101]. Information from narrative text complements the structured data. There are studies that report NLP-processed notes provide more valuable data sources. For example, Penz et al. reports ICD-9 and CPT codes identified less than 11% cases in detecting adverse events related to central venous catheters, while NLP methods achieved a specificity of 0.80 and sensitivity of 0.72 [102]. Widely used general-purpose NLP tools include MedLEE (Medical Language Extraction and Encoding System), cTAKES (clinical Text Analysis and Knowledge Extraction System), MetaMap, and KnowledgeMap. All of them have been successfully applied to phenotyping using EHR data. Task-specific NLP methods are available which aim to extract specific concepts from clinical text.

To build a phenotype algorithm, first we need to select the phenotype of interest, followed by the identification of key clinical elements which 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. Medication record can be used to increase the accuracy of case and control identification of phenotyping algorithms. Patients believed to be controls should be having different medication profile. They may not have any medication prescribed 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.

The DNA sequence of a person can be huge in size, hundreds of gigabytes to terabytes in raw format which exceeds the capability of current EHR systems. Storing, managing and transferring a repository of such large volume of data is difficult. Efficient data compression techniques can be applied to solve this problem. Genome Wide Association Study (GWAS) became the mainstay of genetic analysis over last decade. In general GWAS investigates 500000 genetic variants (Single Nucleotide Polymorphism) or more to see the association of variations with observable traits. It compares the SNPs of cases versus controls to find meaningful knowledge. Besides traits, we can also identify SNPs which determine particular drug response. Some person might react adversely to a particular drug while others might not. Genetic profile of a person can be used for personalized medicine. One big advantage of genetic data is SNPs are same for a person. The same set of data can be used for different phenotype investigation. Researchers are working to integrate genetic information for enhanced clinical decision support. For example, researchers in Vanderbilt University are working on implementing Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment (PREDICT) [103]. St. Jude Children's Research Hospital also has a multiplexed genotyping platform for providing decision support [104].

1.10 Conclusions

Electronic health records are the obvious and inevitable future of patient care in hospitals and medical practices. Widespread adoption of EHRs can foster the improvement of quality in healthcare services, safety and efficiency, and most importantly, public health. Having great potential for benefits, successful deployment of EHRs has several challenges to overcome. There are notable limitations of the use of EHR data in research purposes. In the era of cutting edge technology, the necessary laws lag far behind. While other developed countries have showed widespread adoption, in United States the overall adoption is considerably low. Greater government initiatives and enhanced standardization can lead to brighter healthcare tomorrow.

Bibliography

[1] Eric Jamoom et al. *Physician adoption of electronic health record systems: United States, 2011*, volume 98. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2012.

[2] Mark A Musen and Jan Hendrik Bemmel. *Handbook of medical informatics*. MIEUR, 1999.

[3] Richard S Dick and Elaine B Steen. *The computer-based patient record: an essential technology for health care*. Natl Academy Pr, 1991.

[4] `http://www.healthit.gov/providers-professionals/benefits-electronic-health-records-ehrs`.

[5] George W Bush. Transforming health care: the president's health information technology plan. *Washington, DC*, 2004.

[6] `http://www.cnbc.com/id/28559492/Text_of_Obama_Speech_on_the_Econom`.

[7] `http://www.cdc.gov/osels/lspppo/healthcare_news.html`.

[8] Hagop S Mekhjian, Rajee R Kumar, Lynn Kuehn, Thomas D Bentley, Phyllis Teater, Andrew Thomas, Beth Payne, and Asif Ahmad. Immediate benefits realized following implementation of physician order entry at an academic medical center. *Journal of the American Medical Informatics Association*, 9(5):529–539, 2002.

[9] Virginia McLean. Electronic health records overview. *Center for Enterprise Modernization Mitre. National Institutes of Health National Center for Research Resources*, 2006.

[10] `http://www.who.int/classifications/icd/en/`.

[11] Heinz Katschnig. Are psychiatrists an endangered species? observations on internal and external challenges to the profession. *World Psychiatry*, 9(1):21–28, 2010.

[12] `http://www.who.int/classifications/icd/factsheet/en/index.html`.

[13] `http://www.ncvhs.hhs.gov/031105a1.htm`.

[14] Hude Quan, Bing Li, L Duncan Saunders, Gerry A Parsons, Carolyn I Nilsson, Arif Alibhai, and William A Ghali. Assessing validity of icd-9-cm and icd-10 administrative data in recording clinical conditions in a unique dually coded database. *Health services research*, 43(4):1424–1441, 2008.

[15] `http://www.who.int/classifications/icd/revision/timeline/en/`.

[16] World Health Organization et al. Icd-11 beta draft, 2013.

[17] http://www.who.int/classifications/icd/revision/icd11betaknownconcerns.pdf.

[18] Ronald Cornet and Nicolette de Keizer. Forty years of snomed: a literature review. *BMC medical informatics and decision making*, 8(Suppl 1):S2, 2008.

[19] Michael Q Stearns, Colin Price, Kent A Spackman, and Amy Y Wang. Snomed clinical terms: overview of the development process and project status. In *Proceedings of the AMIA Symposium*, page 662. American Medical Informatics Association, 2001.

[20] http://www.ihtsdo.org/snomed-ct.

[21] Karen Kostick et al. Snomed ct integral part of quality ehr documentation. *Journal of AHIMA/American Health Information Management Association*, 83(10):72–75, 2012.

[22] Clem McDonald, SM Huff, J Suico, and Kathy Mercer. Logical observation identifiers names and codes (loinc r) users' guide. *Indianapolis: Regenstrief Institute*, 2004.

[23] Alex AT Bui and Ricky K Taira. *Medical imaging informatics.*

[24] Casey C Bennett. Utilizing rxnorm to support practical computing applications: Capturing medication history in live electronic health records. *Journal of biomedical informatics*, 45(4):634–641, 2012.

[25] Simon Liu, Wei Ma, Robin Moore, Vikraman Ganesan, and Stuart Nelson. Rxnorm: prescription for electronic drug information exchange. *IT professional*, 7(5):17–23, 2005.

[26] https://www.nlm.nih.gov/research/umls/rxnorm/history.html.

[27] National Center for Health Statistics (US), World Health Organization, et al. *Classification of diseases, functioning, and disability.* US Centers for Disease Control and Prevention.

[28] Alan M Jette. Toward a common language for function, disability, and health. *Journal of American Physical Therapy Association*, 2006.

[29] World Health Organization et al. The who family of international classifications. *Extraído de [http://www. who. int/classifications/en], acesso em [24 de março de 2009]*, 2011.

[30] Helena Hemmingsson and Hans Jonsson. An occupational perspective on the concept of participation in the international classification of functioning, disability and health—some critical remarks. *The American Journal of Occupational Therapy*, 59(5):569–576, 2005.

[31] Rune J Simeonsson, Anita A Scarborough, and Kathleen M Hebbeler. Icf and icd codes provide a standard language of disability in young children. *Journal of Clinical Epidemiology*, 59(4):365–373, 2006.

[32] http://www.ihtsdo.org/snomed-ct/snomed-ct0/.

[33] Norbert Goldfield. The evolution of diagnosis-related groups (drgs): from its beginnings in case-mix and resource use theory, to its implementation for payment and now for its current utilization for quality within and outside the hospital. *Quality Management in Healthcare*, 19(1):3–16, 2010.

[34] WC Hsiao, HM Sapolsky, DL Dunn, and SL Weiner. Lessons of the new jersey drg payment system. *Health Affairs*, 1986.

[35] Catherine Selden and Betsy L Humphreys. *Unified Medical Language System (UMLS): January 1986 Through December 1996: 280 Selected Citations*, volume 96. US Dept. of Health and Human Services, Public Health Service, 1997.

[36] Donald AB Lindberg and Betsy L Humphreys. Concepts, issues, and standards. current status of the nlm's umls project: The umls knowledge sources: Tools for building better user interfaces. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 121. American Medical Informatics Association, 1990.

[37] W Dean Bidgood, Steven C Horii, Fred W Prior, and Donald E Van Syckle. Understanding and using dicom, the data interchange standard for biomedical imaging. *Journal of the American Medical Informatics Association*, 4(3):199–212, 1997.

[38] Asuman Dogac, Tuncay Namli, Alper Okcan, Gokce B Laleci, Yildiray Kabak, and Marco Eichelberg. Key issues of technical interoperability solutions in ehealth and the ride project. *Software R&D Center, Dept. of Computer Eng., Middle East Technical University, Ankara*, 6531, 2007.

[39] J Grabenweger and G Duftschmid. Ontologies and their application in electronic health records. *eHealth2008 Medical Informatics meets eHealth, Wien*, pages 29–30, 2008.

[40] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.

[41] Samuel J Wang, Blackford Middleton, Lisa A Prosser, Christiana G Bardon, Cynthia D Spurr, Patricia J Carchidi, Anne F Kittler, Robert C Goldszer, David G Fairchild, Andrew J Sussman, et al. A cost-benefit analysis of electronic medical records in primary care. *The American journal of medicine*, 114(5):397–403, 2003.

[42] Tricia L Erstad. Analyzing computer-based patient records: A review of literature. *Journal of Healthcare Information Management*, 17(4):51–57, 2003.

[43] Douglas Johnston, Eric Pan, and J Walker. The value of cpoe in ambulatory settings. *J Healthc Inf Manag*, 18(1):5–8, 2004.

[44] Nir Menachemi and Robert G Brooks. Reviewing the benefits and costs of electronic health records and associated patient safety technologies. *Journal of Medical Systems*, 30(3):159–168, 2006.

[45] T Ewing and D Cusick. Knowing what to measure. *Healthcare financial management: journal of the Healthcare Financial Management Association*, 58(6):60–63, 2004.

[46] Abha Agrawal. Return on investment analysis for a computer-based patient record in the outpatient clinic setting. *Journal of the Association for Academic Minority Physicians: the official publication of the Association for Academic Minority Physicians*, 13(3):61–65, 2002.

[47] Robert H Miller, Christopher West, Tiffany Martin Brown, Ida Sim, and Chris Ganchoff. The value of electronic health records in solo or small group practices. *Health Affairs*, 24(5):1127–1137, 2005.

[48] Jeff Mildon and Todd Cohen. Drivers in the electronic medical records market. *Health management technology*, 22(5):14, 2001.

[49] Karl F Schmitt and David A Wofford. Financial analysis projects clear returns from electronic medical records. *Healthcare financial management: journal of the Healthcare Financial Management Association*, 56(1):52–57, 2002.

[50] K Renner. Electronic medical records in the outpatient setting (part 1). *Medical group management journal/MGMA*, 43(3):52–54, 1995.

[51] E Jamoom. National perceptions of ehr adoption: Barriers, impacts, and federal policies. *National Conference on Health Statistics*, 2012.

[52] Hardeep Singh, Traber Davis Giardina, Samuel N Forjuoh, Michael D Reis, Steven Kosmach, Myrna M Khan, and Eric J Thomas. Electronic health record-based surveillance of diagnostic errors in primary care. *BMJ quality & safety*, 21(2):93–100, 2012.

[53] J Michael Fitzmaurice, Karen Adams, and John M Eisenberg. Three decades of research on computer applications in health care medical informatics support at the agency for healthcare research and quality. *Journal of the American Medical Informatics Association*, 9(2):144–160, 2002.

[54] R Lamar Duffy, Shih Shen Angela Yiu, Ehab Molokhia, Robert Walker, and R Allen Perkins. Effects of electronic prescribing on the clinical practice of a family medicine residency. *Fam Med*, 42(5):358–63, 2010.

[55] Linda T Kohn, Janet M Corrigan, Molla S Donaldson, et al. *To err is human: building a safer health system*, volume 627. National Academies Press, 2000.

[56] David W Bates, Lucian L Leape, David J Cullen, Nan Laird, Laura A Petersen, Jonathan M Teich, Elizabeth Burdick, Mairead Hickey, Sharon Kleefield, Brian Shea, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *Jama*, 280(15):1311–1316, 1998.

[57] Ross Koppel, Joshua P Metlay, Abigail Cohen, Brian Abaluck, A Russell Localio, Stephen E Kimmel, and Brian L Strom. Role of computerized physician order entry systems in facilitating medication errors. *Jama*, 293(10):1197–1203, 2005.

[58] Institute of Medicine (US). Committee on Quality of Health Care in America. *Crossing the quality chasm: A new health system for the 21st century.* National Academies Press, 2001.

[59] Elizabeth E Stewart, Paul A Nutting, Benjamin F Crabtree, Kurt C Stange, William L Miller, and Carlos Roberto Jaén. Implementing the patient-centered medical home: observation and description of the national demonstration project. *The Annals of Family Medicine*, 8(Suppl 1):S21–S32, 2010.

[60] Christopher J Stille, Anthony Jerant, Douglas Bell, David Meltzer, and Joann G Elmore. Coordinating care across diseases, settings, and clinicians: a key role for the generalist in practice. *Annals of internal medicine*, 142(8):700–708, 2005.

[61] Gallup serious chronic illness survey 2002. `http://poll.gallup.com/content/default.aspx?ci=6325&pg=1`.

[62] Lynda C Burton, Gerard F Anderson, and Irvin W Kues. Using electronic health records to help coordinate care. *Milbank Quarterly*, 82(3):457–481, 2004.

[63] Ilana Graetz, Mary Reed, Thomas Rundall, Jim Bellows, Richard Brand, and John Hsu. Care coordination and electronic health records: connecting clinicians. In *AMIA Annual Symposium Proceedings*, volume 2009, page 208. American Medical Informatics Association, 2009.

[64] Medicare and medicaid ehr incentive program basics. `http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Basics.html`.

[65] Charles Safran, Meryl Bloomrosen, W Edward Hammond, Steven Labkoff, Suzanne Markel-Fox, Paul C Tang, and Don E Detmer. Toward a national framework for the secondary use of health data: an american medical informatics association white paper. *Journal of the American Medical Informatics Association*, 14(1):1–9, 2007.

[66] Bhanukiran Vinzamuri and Chandan K Reddy. Cox regression with correlation based regularization for electronic health records.

[67] Dipak Kalra and BGME Blobel. Semantic interoperability of ehr systems. *Studies in health technology and informatics*, 127:231, 2007.

[68] Ashish K Jha, David Doolan, Daniel Grandt, Tim Scott, and David W Bates. The use of health information technology in seven nations. *International journal of medical informatics*, 77(12):848–854, 2008.

[69] Elizabeth Davidson and Dan Heslinga. Bridging the it adoption gap for small physician practices: An action research study on electronic health records. *Information Systems Management*, 24(1):15–28, 2006.

[70] Catherine M DesRoches, Eric G Campbell, Sowmya R Rao, Karen Donelan, Timothy G Ferris, Ashish Jha, Rainu Kaushal, Douglas E Levy, Sara Rosenbaum, Alexandra E Shields, et al. Electronic health records in ambulatory care—a national survey of physicians. *New England Journal of Medicine*, 359(1):50–60, 2008.

[71] Ebrahim Randeree. Exploring physician adoption of emrs: a multi-case analysis. *Journal of medical systems*, 31(6):489–496, 2007.

[72] Robert H Miller and Ida Sim. Physicians' use of electronic medical records: barriers and solutions. *Health affairs*, 23(2):116–126, 2004.

[73] Steven R Simon, Rainu Kaushal, Paul D Cleary, Chelsea A Jenter, Lynn A Volk, E John Orav, Elisabeth Burdick, Eric G Poon, and David W Bates. Physicians and electronic health records: a statewide survey. *Archives of Internal Medicine*, 167(5):507–512, 2007.

[74] Glenn A Loomis, J SCOTT Ries, Robert M Saywell, and NITESH R Thakker. If electronic medical records are so great, why aren't family physicians using them? *Journal of Family Practice*, 51(7):636–641, 2002.

[75] Alex R Kemper, Rebecca L Uren, and Sarah J Clark. Adoption of electronic health records in primary care pediatric practices. *Pediatrics*, 118(1):e20–e24, 2006.

[76] 2003 national survey of physicians and quality of care. `http://www.commonwealthfund.org/Surveys/2003/2003-National-Survey-of-Physicians-and-Quality-of-Care.aspx`.

[77] Arun Vishwanath and Susan D Scamurra. Barriers to the adoption of electronic health records: using concept mapping to develop a comprehensive empirical model. *Health Informatics Journal*, 13(2):119–134, 2007.

[78] Nir Menachemi and Taleah H Collum. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy*, 4:47–55, 2011.

[79] Clinicians and ehr usability. `http://www.himss.org/resourcelibrary/TopicList.aspx?MetaDataID=1721&navItemNumber=17121`.

[80] Jeffery L Belden, Rebecca Grayson, and Janey Barnes. Defining and testing emr usability: Principles and proposed methods of emr usability evaluation and rating. Technical report, Healthcare Information and Management Systems Society (HIMSS), 2009.

[81] Ignacio Valdes, David C Kibbe, Greg Tolleson, Mark E Kunik, and Laura A Petersen. Barriers to proliferation of electronic medical records. *Informatics in primary care*, 12(1), 2004.

[82] About cchit. `https://www.cchit.org/about`.

[83] US Department of Health, Human Services, et al. Health information technology: Initial set of standards, implementation specifications, and certification criteria for electronic health record technology. *Federal Register*, 75(8):13, 2010.

[84] Ben Smith, Andrew Austin, Matt Brown, Jason T King, Jerrod Lankford, Andrew Meneely, and Laurie Williams. Challenges for protecting the privacy of health information: required certification can leave common vulnerabilities undetected. In *Proceedings of the second annual workshop on Security and privacy in medical and home-care systems*, pages 1–12. ACM, 2010.

[85] Sandeep S Mangalmurti, Lindsey Murtagh, and Michelle M Mello. Medical malpractice liability in the age of electronic health records. *New England Journal of Medicine*, 363(21):2060–2067, 2010.

[86] K Bruce Bayley, Tom Belnap, Lucy Savitz, Andrew L Masica, Nilay Shah, and Neil S Fleming. Challenges in using electronic health record data for cer: Experience of 4 learning organizations and solutions applied. *Medical care*, 51:S80–S86, 2013.

[87] Jionglin Wu, Jason Roy, and Walter F Stewart. Prediction modeling using ehr data: challenges, strategies, and a comparison of machine learning approaches. *Medical care*, 48(6):S106–S113, 2010.

[88] Chris Paxton, Alexandru Niculescu-Mizil, and Suchi Saria. Developing predictive models using electronic medical records: Challenges and pitfalls.

[89] Brian J Wells, Amy S Nowacki, Kevin Chagin, and Michael W Kattan. Strategies for handling missing data in electronic health record derived data. *Strategies*, 12:17–2013.

[90] Harvey J Murff, Fern FitzHenry, Michael E Matheny, Nancy Gentry, Kristen L Kotter, Kimberly Crimin, Robert S Dittus, Amy K Rosen, Peter L Elkin, Steven H Brown, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA*, 306(8):848–855, 2011.

[91] David Blumenthal and Marilyn Tavenner. The "meaningful use" regulation for electronic health records. *New England Journal of Medicine*, 363(6):501–504, 2010.

[92] Shawn N Murphy, Griffin Weber, Michael Mendis, Vivian Gainer, Henry C Chueh, Susanne Churchill, and Isaac Kohane. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *Journal of the American Medical Informatics Association*, 17(2):124–130, 2010.

[93] John F Hurdle, Stephen C Haroldsen, Andrew Hammer, Cindy Spigle, Alison M Fraser, Geraldine P Mineau, and Samir J Courdy. Identifying clinical/translational research cohorts: ascertainment via querying an integrated multi-source database. *Journal of the American Medical Informatics Association*, 20(1):164–171, 2013.

[94] Dan M Roden, Jill M Pulley, Melissa A Basford, Gordon R Bernard, Ellen W Clayton, Jeffrey R Balser, and Dan R Masys. Development of a large-scale de-identified dna biobank to enable personalized medicine. *Clinical Pharmacology & Therapeutics*, 84(3):362–369, 2008.

[95] Catherine A McCarty, Anuradha Nair, Diane M Austin, and Philip F Giampietro. Informed consent and subject motivation to participate in a large, population-based genomics study: the marshfield clinic personalized medicine research project. *Public Health Genomics*, 10(1):2–9, 2006.

[96] P. L. Elkin, A. P. Ruggieri, S. H. Brown, B. A. Bauer, D Wahner-Roedler, S. C. Litin, J Beinborn, K. R. Bailey, and L. Bergstrom. A randomized controlled trial of the accuracy of clinical record retrieval using snomed-rt as compared with icd-9-cm. *American Medical Informatics Association*, 2001.

[97] Joshua C Denny. Mining electronic health records in the genomics era. *PLoS computational biology*, 8(12):e1002823, 2012.

[98] Joshua C Denny, Josh F Peterson, Neesha N Choma, Hua Xu, Randolph A Miller, Lisa Bastarache, and Neeraja B Peterson. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. *Journal of the American Medical Informatics Association*, 17(4):383–388, 2010.

[99] Joshua C Denny, Randolph A Miller Anderson Spickard III, Jonathan Schildcrout, Dawood Darbar, S Trent Rosenbloom, and Josh F Peterson. Identifying umls concepts from ecg impressions using knowledgemap. In *AMIA Annual Symposium Proceedings*, volume 2005, page 196. American Medical Informatics Association, 2005.

[100] Elena Birman-Deych, Amy D Waterman, Yan Yan, David S Nilasena, Martha J Radford, and Brian F Gage. Accuracy of icd-9-cm codes for identifying cardiovascular and stroke risk factors. *Medical care*, 43(5):480–485, 2005.

[101] Elizabeth FO Kern, Miriam Maney, Donald R Miller, Chin-Lin Tseng, Anjali Tiwari, Mangala Rajan, David Aron, and Leonard Pogach. Failure of icd-9-cm codes to identify patients with comorbid chronic kidney disease in diabetes. *Health services research*, 41(2):564–580, 2006.

[102] Janet FE Penz, Adam B Wilcox, and John F Hurdle. Automated identification of adverse events related to central venous catheters. *Journal of biomedical informatics*, 40(2):174–182, 2007.

[103] Jill M Pulley, Joshua C Denny, Josh F Peterson, Gordon R Bernard, Cindy L Vnencak-Jones, Andrea H Ramirez, Jessica T Delaney, Erica Bowton, Kyle Brothers, Kevin Johnson, et al. Operational implementation of prospective genotyping for personalized medicine: the design of the vanderbilt predict project. *Clinical Pharmacology & Therapeutics*, 92(1):87–95, 2012.

[104] J Kevin Hicks, Kristine R Crews, James M Hoffman, Nancy M Kornegay, Mark R Wilkinson, Rachel Lorier, Alex Stoddard, Wenjian Yang, Colton Smith, Christian A

Fernandez, et al. A clinician-driven automated system for integration of pharmacogenetic consults into an electronic medical record. *Clinical pharmacology and therapeutics*, 92(5):563, 2012.