Development of Hospital-based Data Sets as a Vehicle for Implementation of a National Electronic Health Record

by Leila Keikha, MSc; Seyede Sedigheh Seied Farajollah, MSc; Reza Safdari, PhD; Marjan Ghazisaeedi, PhD; and Niloofar Mohammadzadeh, PhD

Abstract

Background: In developing countries such as Iran, international standards offer good sources to survey and use for appropriate planning in the domain of electronic health records (EHRs). Therefore, in this study, HL7 and ASTM standards were considered as the main sources from which to extract EHR data.

Objective: The objective of this study was to propose a hospital data set for a national EHR consisting of data classes and data elements by adjusting data sets extracted from the standards and paper-based records.

Method: This comparative study was carried out in 2017 by studying the contents of the paper-based records approved by the health ministry in Iran and the international ASTM and HL7 standards in order to extract a minimum hospital data set for a national EHR.

Results: As a result of studying the standards and paper-based records, a total of 526 data elements in 174 classes were extracted. An examination of the data indicated that the highest number of extracted data came from the free text elements, both in the paper-based records and in the standards related to the administrative data. The major sources of data extracted from ASTM and HL7 were the E1384 and Hl7V.x standards, respectively. In the paper-based records, data were extracted from 19 forms sporadically.

Discussion: By declaring the confidentiality of information, the ASTM standards acknowledge the issue of confidentiality of information as one of the main challenges of EHR development, and propose new types of admission, such as teleconference, tele-video, and home visit, which are inevitable with the advent of new technology for providing healthcare and treating diseases. Data related to finance and insurance, which were scattered in different categories by three organizations, emerged as the financial category. Documenting the role and responsibility of the provider by adding the authenticator/signature data element was deemed essential.
Conclusion: Not only using well-defined and standardized data, but also adapting EHR systems to the local facilities and the existing social and cultural conditions, will facilitate the development of structured data sets.

Keywords: hospital data; data set; electronic health record; comparative study

Introduction

In the different electronic health record (EHR) systems available on the market, health data are stored in different formats, which not only poses a difficulty for the sellers, buyers, and users of these systems but also complicates the exchange of data.[1] Healthcare providers need a well-defined and standardized data set so that they can exchange data in a standard format and transfer health data from one center to another. This way, the goal of the integrity of the EHR is achieved.[2] In fact, we can prevent the creation and spread of idiosyncratic information systems by establishing a standard data set accepted at the national level.[3] The minimum data set is the minimal necessary number of the main variables related to the individual's health status and the patient's care plan[4] within the hospital data set in this study. The hospital variables related to the individual conditions include demographic, financial, clinical, and care plan data.[5–7] Previous studies have indicated that standards are good sources for obtaining appropriate data elements for the EHR.[8–10]

Several organizations all over the world are striving to coordinate and promote EHR standards. As international organizations pioneering in the development of EHR standards, HL7 and ASTM organizations were appropriate sources from which to extract EHR data elements.[11–16] In these sources, the data are presented in two formats: structured and free text. In different countries, during the stages of codifying a proper structure and content for the EHR, the suitable data sets were codified. However, each country applied different methods for identification of the data elements and their placement within the minimum data set content. For example, in China, Tu et al. developed an electronic medical record (EMR) data set that was based on the paper-based medical records of the Chinese hospitals and designed along the HL7 CDA architecture.[17] In 2011, a minimum clinical cardiovascular data set was codified as part of a collaboration by the American College of Cardiology and a task force of the American Heart Association (AHA), in which the ISO/IEC 11179 standard and a survey of experts were used as a basis for the design.[18] In a study conducted by Xu et al. in China, the ASTM E1384 and ISO18308 standards were used for the evaluation of EHR contents.[19] Watzlaf et al.[20] utilized ASTM E1384 for the evaluation of EHR data sets.

In this paper, a minimum hospital data set for a national EHR was proposed by adjusting the data sets extracted from standards and paper-based records.

Methods

In this descriptive study conducted in 2017, development of a hospital data set was done in five steps as follows:

In the first step, in order to review the texts, the website and manuals of each organization were consulted, and correspondence was established with the organizations' managers and authorities. The relevant articles published in the English language were extracted, without any time limitation, by searching scientific and authoritative databases such as ScienceDirect, ProQuest, Scopus, Google Scholar, and PubMed for keywords including *minimum data set*, *EHR*, *HL7*, and *ASTM*. The studies that examined the concept of a minimum data set and/or its evaluation in the EHR were included in this study. On the basis of these studies, the E1384, E1238, E1239, E2473, E1238, E1239, and E2084 standards were chosen as the sources for ASTM data extraction. The HL7V.X standard and PD1, ADT, GT1, MPI, DRG, ACC, PV1, and PV2 messages were the sources for extraction of HL7 data.

To extract the data from paper records, the main sheets of the paper-based records approved by the Ministry of Health in Iran were studied. These sheets included the following: (1) admission and discharge summary, (2) unit summary, (3) medical history and physical exam, (4) progress note, (5) physician's order, (7) nurses' note, (8) anesthesia, (9) preoperative care, (10) operation report, (11) postoperative care, (12) vital signs control, (13) vital signs chart, (14) laboratory report attachment, (15) electrocardiogram attachment, (16) pathology report, (17) radiology report, (18) fluid balance chart, and (19) consultation request.

It should be noted that the specialized sheets, such as the burn sheet and the hemodialysis and peritoneal dialysis sheets, that are incorporated into the records of patients with certain diseases, were not taken into consideration in this study. Additionally, the codes related to the patient's disease and condition, which exist in the unit summary sheet of the paper-based records and are presented separately in the standards framework and terminology, were not considered in this paper.

In the next step, to extract the data elements, we surveyed the standards and the paper records. In some sources, such as ASTME1384, the content of the EHR was extracted directly, whereas it was necessary to survey the content of other standards to extract related data, such as HL7 messages. Therefore, the authors identified the data elements in face-to-face meetings after surveying the standards.

Next, the data elements obtained from the ASTM and HL7 organizations were adjusted to correspond with the paper-based records. In this step, the granularity (level of detail) of the data, the presence or lack of data in each category, and the extent of the relationship with the intended category were considered as the areas for comparison in this study.

The data elements were categorized according to a checklist provided in 2006 by AHIMA to develop a minimum data set for EHRs. This checklist includes 10 categories; administrative, encounter, problem, treatment plan, provider, evaluation, diagnostic tests, history, event, and insurance services.

Finally, the extracted data were divided into classes (headings) and data elements (subheadings).

Results

As a result of the study, 526 data elements in 174 classes in 10 categories were extracted. The format of most of the data was free text. Data classes were considered as headings, and generally each class included data elements as subheadings. The numbers of the classes and data elements extracted from the standards and paper-based records are presented in Table 1. In addition, the different sources of extracted data, including standard names and paper record sheets, are stated in Table 2.

Data can be recorded in the hospital records in both structured and free-text formats. Structured data elements are determined with a distinct value, such as terminology codes or numbered values,[21] whereas free-text data are recorded as text or narratives.[22, 23] Furthermore, in the ASTM standards, some class data and data elements are also included within the main tables' framework.[24] These tables include a list of variables specifying the attributes of an object.[25, 26] For example, in the administrative data, the inpatient reception, the outpatient reception, and the diagnosis upon reception are presented in the main tables. In the encounter data, the name of the problem in each encounter and the diagnosis at encounter are provided in the main tables. In Table 3, the data element categories are presented in terms of their structured or free-text format.

Category 1: Administrative Data

The administrative data in this category include personal data elements for identifying the patient's disease. These items are obtained from the patient or the patient's legal representative. Some of these elements are repeated from encounter to encounter, and others need to be updated in each encounter.[27, 28] In this study, 102 administrative data elements were extracted (more details are provided in Table 1).

The comparison of the data items extracted from the organizations revealed that the greatest number of extracted data items was associated with the administrative data, both in the standards and in the paper-based record sheets, but the standards addressed these data differently and in greater detail. For instance, the insurance identification number and the driver's license number are presented in the standards but are not provided in the paper record sheets. Furthermore, the standards included an "unspecified" option for the sex data and presented the marital status data in more detail. Some data elements, such as race and mother's full name, were mentioned in only one standard. The ASTM standards presented the category of data confidentiality with very limited control, limited, and normal control details, thereby taking the issue of information security into account. ASTM provides information regarding the record identifier (ID), including the inclusive ID, in the header section and presents the insurance information in the body section.[29] For further information on the sources of extracted data, see in Table 2.

After comparing the data, we determined that the main categories of administrative data included patient name, identifier number, sex, religion, birth data, father's full name, mother's full name, consent, patient's language, education level, marital status, type of confidentiality, occupation, date of registration, ethnic groups, address, person who brought or referred the patient, person the encounter was with, blood group, Rh type, directive to physician, medicolegal case type, and additional demographic data.

Category 2: Encounter Data

The encounter is defined as a face-to-face session with the physician or therapy center, during which the administrative or clinical information is exchanged. More detail on the encounter data category, with 128 data elements, is provided in Table 1.

The categorization of the encounter type is considered in the admission and discharge summary and unit summary sheets of the paper-based records. In ASTM standard E1238, new types of encounters such as home visit, tele-video, and teleconference are mentioned.[30] The patient ID data is presented only in E1238, and the discharge data are referred to in HL7 in detail. The place of treatment, which is placed in the header of all the paper-based record sheets, includes a wide range of mobile outpatient centers, health centers, specialized clinics, and nursing care centers in the E1384 standard.[31] Data related to the reception and discharge were incorporated in the header of all the paper record sheets. The sources of extracted data by organizational division are stated in Table 2.

After comparing the data, we determined that the main classes of encounter data included patient ID, encounter type, place of care, admission, history of hospitalization, hospitalization data, differential diagnosis, note/report text, consultation data, length of stay, method of treatment of the patient, death data, and authentication/signature.

Category 3: Problem Data

This section includes clinical problems, diagnosis and exposure risk, a list of health events, the diagnosis of current and previous pathophysiological status, physical symptoms, risk factors, allergies, drug or food reactions, and other behavioral problems.

These data are only presented separately in the ASTM E1238 standard, in the six main classes of problem number, problem ID, problem at encounter, problem at care, problem in current status, and date of problem.

Category 4: Treatment Plan Data

The patient's treatment data, which are recorded through direct or indirect observation, are included in this category. Table 1 shows that 95 data elements were extracted in this category.

The major data elements related to this category involve the data associated with the general and clinical orders. In the paper-based records, the orders are included in the physician's order sheet. In the standards, the orders are presented in two general and clinical sections and are related to the place of treatment and the provider of the treatment prescription. The source of the data elements extracted are presented in Table 2.

After comparing the data, we determined that the main classes of treatment plan data were as follows: general order, observation order, treatment ID, radiology orders and results, diet orders, clinical orders, and pharmacy orders.

Category 5: Provider Data

The provider is a working entity who provides the healthcare for the consumer of services and who is permitted to work in a health center because of their professional expertise in providing healthcare. Table 1 shows that 25 data elements were extracted in this category.

The data related to the provider role, with such details as the identification number, the role of the care provider, the address, and the phone number, are presented in the ASTM standards and the paper-based record sheets. In the paper record, the signature of the provider was included in the related form; for

example, the operation report and the nurses' note were signed by the surgeon and the nurse, respectively. The sources of data are depicted in Table 2.

After comparing the data, we determined that the main classes of provider data included the provider, provider agency ID code, practitioner's current role, practitioner's address, practitioner's phone number, and practitioner's license number.

Category 6: Examination Data

Data on examinations and clinical trials, physician's consultations, and assessments before, during, and after surgery are located in this category. After reviewing the sources, we extracted 110 data elements in this category. Information about the data elements is represented in Table 1.

In the ASTM standards, features such as the ID, the date, and the place are presented. In the HL7 standard, the focus of data is on the laboratory samples and findings. In the paper-based records, the examination data include the physical examination of the body organs and the time of the evaluation, such as preoperative or postoperative. Furthermore, in the ASTM CCRV.1 standard, sections 5 and 6 indicate the patient's health status and care documentation, and the CCRV.1a body section presents the evaluation findings and the results, which include the patient's status, daily activities, mental condition, life condition at home, and personal care capability. The sources of the extracted examination data are stated in Table 2.

The clinical examination, examination during surgery, preoperative examination, postoperative examination, and exam findings were main classes in this category.

Category 7: History Data

These data include allergies, consumed drugs, and health risk factors associated with the patient's condition at home and in society. Fourteen elements in seven classes were obtained from the ASTM standards, and six classes were obtained from the paper-based records. Information about the extracted data items is provided in Table 1.

In the ASTM standards, details of the patient's history data are provided in CCRV.1 in section 5 and CCRV.1a in the body section. In the HL7, data related to the case history did not appear in any standards or messages. Such data were extracted from the medical history and physical exam sheet in the paper-based records.

After comparing the three sources, we determined that this category included the following classes: the source of history name, the health history, the family history, the current drug therapy and other addictions, allergies, the presenting symptoms, the social background history, the past disease history, current habits, and relevant events in the patient's history.

Category 8: Diagnostic Data

Diagnostic data are the details that guide the physician in the diagnosis, management, and treatment of the patient, such as laboratory, radiology, and nuclear medicine findings and other diagnostic evaluations. In overall, 16 data elements were extracted in this category and are shown in detail in Table 1.

The focus of both the standards and the paper-based records is on the pathology, laboratory, and radiology findings. However, no diagnostic data were observed in the standards and the evaluated messages of HL7. The sources of the data are depicted in Table 2.

The main classes of diagnostic data were micro-organism attributes, test request order, treatment facility, test request per sheet facility, micro-organism resistance pattern, micro-organism specifications, micro-organism test request, laboratory, and radiology.

Category 9: Event Data

The category of event data includes the data related to a care event involving the patient. Extracted data elements related to event data are shown in Table 1.

In sections 5 and 6 of ASTM CCRV.1, in which the health status of the patient and the care documentation are outlined, the event data, which has one data element, are considered in four classes. They include the cause of the event, the chief complaint, clinical progress, and the approval of the recorder. In HL7, the messages were the source of the data elements extracted in three classes, which included the event and the visit information.[32] In the paper-based records, these data were extracted from the admission and discharge summary and from unit summary in two classes, which involved the chief complaint and the clinical progress.

Finally, we determined that the main classes of event data were the reason for visit, chief complaint, patient visit information, patient visit additional information, clinical progress note, accident information, and authenticator/signature.

Category 10: Insurance Data

Insurance data consists of all data related to anesthesia, operations, vaccinations, treatment, medicines, and insurance. Data elements extracted in this category are detailed in Table 1.

In the ASTM standards, the data needed to pay the insurance include the postoperative diagnosis, the name of the surgery, the anesthetic agent, the name of the treatment, the medical instrument, and the name of the drug, which are presented as the main tables in the E1384 standard in five main categories and 20 data elements.[33] In HL7, the insurance data were obtained from the HL7V2.5 UB sheets and categories in seven main categories. In the ASTM standards, insurance data were obtained from the DRG code. In the paper-based records, seven classes and 15 data elements were extracted from the reception and discharge summary, anesthesia, preoperative care, surgery report, and postoperative care sheets. The sources of the extracted data items are represented in Table 2.

Insurance data consists of all data that is taken into account for the payment of patient care services in the hospital. As a result, we determined that the main classes in this category are operations, therapies, surgeon, specimen, drug and accident insurance, vaccination, and special care.

Discussion

According to the finding of this study as shown in Table 2, ASTM E1384 and HL7V.X were the major sources of data for extraction. For instance, ASTM E1384 had data in the most categories, such as administrative, encounter, treatment plan, provider, diagnosis, and insurance data. Research conducted in other countries supports this claim. In a study by Watzlaf et al., the E1384 standard was used to measure the minimum validity of the EHR content, and the results of the survey revealed that the majority of the respondents believed that the minimum data elements of this standard should be taken into consideration in the EHR system.[34] In this regard, Xu et al. used the E1384 and ISO18308 standards to evaluate EHR content.[35] In 2011, the American College of Cardiology applied the ISOIEC11179 standard to evaluate the content of a clinical cardiovascular data set.[36] Subsequently, Tu et al. used the HL7 CDA architecture to develop data sets that can be shared among EMR information systems.[37]

The messages related to the exchange of data in the HL7 standards were the main source of clinical data extraction, whereas the admission and discharge summary sheet and unit summary sheet were mainly used in the paper-based records. In the extracted data set, the highest and the lowest numbers of items were associated with the ASTM and the paper-based data, respectively, while HL7 had no data elements in the problem, diagnosis, and history categories. Most of the data were in free-text format, and the ASTM standards presented some data variables within main tables. Therefore, because some categories such as reception type or place of treatment contain numerous variables, the use of this format for categorization of data will facilitate the creation of structured data in the national EHR.

The proposed data set can be divided into three general categories. The first category is the data that are contained in the standards but not in the paper forms, including mother's full name, type of confidentiality, and directive to physician in the administrative category; method of treating the patient in the encounter data; general orders, pathology and radiology, clinical orders, pharmacy, observation, and diet orders in the treatment plan; history- taking event in the history data category; and vaccination,

special care, and accident insurance in the insurance data category. These items were considered for inclusion in the proposed data set. The second group of data was scattered in the paper forms but was categorized in the standards. In this case, we used the AHIMA checklist categories to properly organize the data into diagnostic, event, and provider categories. The third category consisted of data that were documented more comprehensively in the paper form than in the standards, including examination data.

Given that one of the main challenges of EHR implementation is the issue of security and confidentiality of information, we recommend that the confidentiality category presented in ASTM E1384, with very limited, limited, and normal control data elements, should be considered for the optimal use of the EHR. Teleconference, tele-video, and home visits have been taken into account as the new aspects of the different types of encounter. Therefore, reflecting the advent of technology in healthcare, we recommend that the encounter type be considered in the main category. Furthermore, the place of treatment category appearing in the header of all the paper-based sheets, indicating the health center where the treatment occurs, should be taken into account in the EHR because care is provided in a wide range of locations, from rural healthcare centers to urban medical centers as well as social care and rehabilitation centers. To document the role and responsibility of the service providers, adding the authenticator/signature data element to each general category of data is deemed essential.

Because of the overlapping of problem data with encounter data, we advise that the problem data be presented as in the E1384 standard, which reflects the patient's status and the problems included in the encounter data category. Another issue with a noticeable difference between the standards and the paper-based records was the issue of the payment documentation and insurance. In the ASTM standards, the different kinds of insurance, including Medicare and Medicaid, are specified, and the DRG code was considered for the payments. In HL7, universal bill sheets are utilized for the insurance data, whereas in the paper-based records, different financial codes are used, depending on the location of the health center. Hence, it is recommended that all the payment and insurance data be placed in a separate financial data category to ensure the integrity and completeness of the patients' payment data.

Conclusion

Health data have constantly been the basis of decision making and policy making within the healthcare industry and are thought to offer a solution for the challenges associated with insurance issues and financial constraints. Concurrently, interactive health information systems are being developed with the aim of being user-friendly and saving time. Therefore, the development of standard and structured data sets, in accordance with the health facilities and the social and cultural conditions of the society, is one of the most important strides that can be taken to achieve these objectives. In addition to the participation of clinical specialists, health information management and medical informatics professionals, as the principal users of the electronic health system, will facilitate its implementation. Certainly, a future survey of the beneficiaries of the system will make the proposed data set more applicable and user-friendly. It can also be considered as a valuable source of the data elements to be contained in national EHR systems. In this context, we used international standards and the AHIMA checklist approach to prepare a minimum data set for use in the national EHR of Iran, but it can be further developed for use on the international level by studying additional resources, such as European and other international standards. Therefore, the developed data set has the potential to be used internationally. Finally, the authors plan to conduct a survey evaluating the proposed data set among health information management professionals, health informatics specialists, clinical specialists, and the senior managers of healthcare organizations for future publication.

Leila Keikha, MSc, is a PhD student of health information management at Tehran University of Medical Sciences in Tehran, Iran.

Seyede Sedigheh Seied Farajollah, MSc, is a PhD student of health information management at Tehran University of Medical Sciences in Tehran, Iran.

Reza Safdari, PhD, is a professor in the Department of Health Information Management at Tehran University of Medical Sciences in Tehran, Iran.

Marjan Ghazisaeedi, PhD, is an assistant professor in the Department of Health Information Management at Tehran University of Medical Sciences in Tehran, Iran.

Niloofar Mohammadzadeh, PhD, is an assistant professor in the Department of Health Information Management at Tehran University of Medical Sciences in Tehran, Iran.

Notes

1. Eggebraaten, T. J., J. W. Tenner, and .JDubbels C. "A Health-Care Data Model Based on the HL7 Reference Information Model." *IBM Systems Journal* 46, no. 1 (2007): 5–18.
2. Spisla, C. "Enhancement of Interoperability of Disaster-related Data Collection Using Disaster Nursing Minimum Data Set." *Studies in Health Technology and Informatics* 146 (2009): 780–81.
3. Ahmadi, M., J. Alipour, A. Mohammadi, and F. Khorami. "Development a Minimum Data Set of the Information Management System for Burns." *Burns* 41, no. 5 (2015): 1092–99.
4. Cai, S., D. B. Mukamel, P. Veazie, and H. Temkin-Greener. "Validation of the Minimum Data Set in Identifying Hospitalization Events and Payment Source." *Journal of the American Medical Directors Association* 12, no. 1 (2011): 38–43.
5. Kowal, P. R., L. J. Wolfson, and J. E. Dowd. "Creating a Minimum Data Set on Ageing in Sub-Saharan Africa." *Southern African Journal of Gerontology* 9 (2000): 18–23.
6. Watzlaf, V. J., X. Zeng, C. Jarymowycz, and P. A. Firouzan. "Standards for the Content of the Electronic Health Record." *Perspectives in Health Information Management* 1 (2004).
7. Renner, A. L., and J. C. Swart. "Patient Core Data Set: Standard for a Longitudinal Health/Medical Record." *Computers in Nursing* 15, no. 2 (1996): S7–S13.
8. Ibid.
9. Watzlaf, V. J., X. Zeng, C. Jarymowycz, and P. A. Firouzan. "Standards for the Content of the Electronic Health Record."
10. Kwak, Y. S. "International Standards for Building Electronic Health Record (EHR)." *Proceedings of 7th International Workshop on Enterprise Networking and Computing in Healthcare Industry, 2005 HEALTHCOM* (2005): 18–23.
11. Mon, D. T. "HL7 EHR System Functional Model and Standard." Presented at the HIMSS Annual Conference, February 2011.
12. Garde, S., P. Knaup, E. J. Hovenga, and S. Heard. "Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for Open EHR Archetypes." *Methods of Information in Medicine* 46, no. 3 (2007): 332–43.
13. Schloeffel, P., T. Beale, G. Hayworth, S. Heard, and H. Leslie. "The Relationship between CEN 13606, HL7, and OpenEHR." *HIC 2006 and HINZ 2006: Proceedings* (2006): 24.
14. Orlova, A. O., M. Dunnagan, T. Finitzo, M. Higgins, T. Watkins, A. Tien, and S. Beales. "An Electronic Health Record-Public Health (EHR-PH) System Prototype for Interoperability in 21st Century Healthcare Systems." *AMIA Annual Symposium Proceedings* (2005): 575–79.
15. Garets, D., and M. Davis. "Electronic Patient Records." *Healthcare Informatics Online* 4 (2005).
16. Jian, W. S., C. Y. Hsu, T. H. Hao, H. C. Wen, M. H. Hsu, Y. L. Lee, et al. "Building a Portable Data and Information Interoperability Infrastructure—Framework for a Standard Taiwan Electronic Medical Record Template." *Computer Methods and Programs in Biomedicine* 88, no. 2 (2007): 102–11.
17. Tu, H., Y. Yu, P. Yang, X. Tang, J. Hu, K. Rao, et al. "Building Clinical Data Groups for Electronic Medical Record in China." *Journal of Medical Systems* 36, no. 2 (2012): 723–36.

18   Weintraub, W. S., R. P. Karlsberg, J. E. Tcheng, J. R. Boris, A. E. Buxton, J. T. Dove, et al. "ACCF/AHA 2011 Key Data Elements and Definitions of a Base Cardiovascular Vocabulary for Electronic Health Records." *Circulation* 124, no. 1 (2011): 103–23.

19.   Xu, W., Z. Guan, H. Cao, H. Zhang, M. Lu, and T. Li. "Analysis and Evaluation of the Electronic Health Record Standard in China: A Comparison with the American National Standard ASTM E 1384." *International Journal of Medical Informatics* 80, no. 8 (2011): 555–61.

20.   Watzlaf, V. J., X. Zeng, C. Jarymowycz, and P. A. Firouzan. "Standards for the Content of the Electronic Health Record."

21.   Baron, R. J. "Quality Improvement with an Electronic Health Record: Achievable, But Not Automatic." *Annals of Internal Medicine* 147, no. 8 (2007): 549–52.

22.   Rosenbloom, S. T., J. C. Denny, H. Xu, N. Lorenzi, W. W. Stead, and K. B. Johnson. "Data from Clinical Notes: A Perspective on the Tension between Structure and Flexible Documentation." *Journal of the American Medical Informatics Association* 18, no. 2 (2011): 181–86.

23.   Rea, S., J. Pathak, G. Savova, T. A. Oniki, L. Westberg, C. E. Beebe, et al. "Building a Robust, Scalable and Standards-driven Infrastructure for Secondary Use of EHR Data: The SHARPn Project." *Journal of Biomedical Informatics* 45, no. 4 (2012): 763–71.

24.   Xu, W., Z. Guan, H. Cao, H. Zhang, M. Lu, and T. Li. "Analysis and Evaluation of the Electronic Health Record Standard in China: A Comparison with the American National Standard ASTM E 1384."

25.   Watzlaf, V. J., X. Zeng, C. Jarymowycz, and P. A. Firouzan. "Standards for the Content of the Electronic Health Record."

26.   Iqbal, A. M., M. Shepherd, and S. S. R. Abidi. "An Ontology-based Electronic Medical Record for Chronic Disease Management." Presented at the IEEE 44th Hawaii International Conference on System Sciences (HICSS), 2011.

27.   Pasquali, S. K., J. P. Jacobs, G. J. Shook, S. M. O'Brien, M. Hall, M. L. Jacobs, et al. "Linking Clinical Registry Data with Administrative Data Using Indirect Identifiers: Implementation and Validation in the Congenital Heart Surgery Population." *American Heart Journal* 160, no. 6 (2010): 1099–1104.

28.   Peabody, J. W., J. Luck, S. Jain, D. Bertenthal, and P. Glassman. "Assessing the Accuracy of Administrative Data in Health Information Systems." *Medical Care* 42, no. 11 (2004): 1066–72.

29.   ASTM International. ASTM E2369-12, Standard Specification for Continuity of Care Record (CCR). Available at https://www.astm.org/Standards/E2369.htm.

30.   ASTM International. ASTM E1238-97, Standard Specification for Transferring Clinical Observations Between Independent Computer Systems. Available at https://www.astm.org/DATABASE.CART/WITHDRAWN/E1238.htm.

31.   ASTM International. ASTM E1384-07(2013), Standard Practice for Content and Structure of the Electronic Health Record (EHR). Available at https://www.astm.org/Standards/E1384.htm.

32.   HL7 International. Available at http://hl7-definition.caristix.com:9010/Default.aspx?version=HL7+v2.5.1&segment=PV1

33.   ASTM International. ASTM E1384-07(2013), Standard Practice for Content and Structure of the Electronic Health Record (EHR).

34.   Watzlaf, V. J., X. Zeng, C. Jarymowycz, and P. A. Firouzan. "Standards for the Content of the Electronic Health Record."

35. Xu, W., Z. Guan, H. Cao, H. Zhang, M. Lu, and T. Li. "Analysis and Evaluation of the Electronic Health Record Standard in China: A Comparison with the American National Standard ASTM E 1384."

36. Weintraub, W. S., R. P. Karlsberg, J. E. Tcheng, J. R. Boris, A. E. Buxton, J. T. Dove, et al. "ACCF/AHA 2011 Key Data Elements and Definitions of a Base Cardiovascular Vocabulary for Electronic Health Records."

37. Tu, H., Y. Yu, P. Yang, X. Tang, J. Hu, K. Rao, et al. "Building Clinical Data Groups for Electronic Medical Record in China."

Table 1

Frequency of Classes and Data Elements Extracted from Standards and Paper-based Record Sheets

| Source | Data Categories | | | | | | | | | | | | | | | | | | | |
| | Administrative | | Encounter | | Problem | | Treatment Plan | | Provider | | Exam | | Diagnosis | | History | | Event | | Insurance | |
	CL	DE	CL	DE	CL	DE	CL	DE	CL	DE	CL	DE	CL	DE	CL	DE	CL	DE	CL	DE
ASTM	23	48	10	51	6	–	7	68	1	7	4	13	11	–	7	14	4	1	5	20
HL7	23	34	3	10	–	–	3	24	1	–	1	6	–	–	–	–	3	–	8	–
Paper record	12	20	12	67	-	–	1	3	3	18	8	91	3	16	6	–	2	–	7	15
Total	58	102	25	128	6	–	11	95	5	25	13	110	14	16	13	14	9	1	20	35

Abbreviations: CL, class; DE, data element.

Table 2

Sources of the Data Extracted from the Standards and Paper-based Record Sheets

Source	Administrative	Encounter	Problem	Treatment Plan	Provider	Exam	Diagnosis	History	Event	Insurance
ASTM	E1238, E1239[1], E1384, E2473[2], and CCRV.1a,	E1238,E1239, E1384, E2084[3]	E1238	E1384, CCRV.1, CCRV.1a	E1384, CCRV.1, CCRV.1a	CCRV.1, CCRV.1a	CCRV.1, CCRV.1a	E1384, CCRV.1	CCRV.1	E1384
HL7	HL7v.x standard,[a] and the messages[b] and the functions of EHR-DSTU[c]	HL7V.X standard[a] and the messages[b]	--	HL7V.X	HL7V.X	HL7V.X	-----	-------	Messages	HL7V2.5, UB sheets
Paper record	Identification data in all the sheets as header, other data in the reception sheets, the discharge summary, the record summary, anesthesia form, and preoperative care form	Admission and discharge summary, unit summary, progress note, physician' order, nurses' note, anesthesia, preoperative care, operation report, postoperative care, vital signs chart, laboratory report attachment, pathology report, radiology report, consultation request, the electrocardiogram	-----	Physician's orders, postoperative care, and nurses' note	All related sheets	Admission and discharge summary, medical history and physical exam, anesthesia, preoperative care, operation report, postoperative care, vital signs control, vital signs chart, fluid balance chart and ,consultation request sheet	Medical history and physical exam	Admission and discharge summary sheet, pathology report, radiology report, laboratory attachment report, and the electrocardiogram attachment sheets	Admission and discharge summary, unit summary	Reception sheet and the discharge summary, anesthesia, preoperative care, the operation report, and postoperative care sheets

Sources:

1. ASTM International. ASTM E1239-04(2010), Standard Practice for Description of Reservation/Registration-Admission, Discharge, Transfer (R-ADT) Systems for Electronic Health Record (EHR) Systems. Available at https://www.astm.org/Standards/E1239.htm.
2. ASTM International. ASTM E2473-05(2011), Standard Practice for the Occupational/Environmental Health View of the Electronic Health Record. Available at https://www.astm.org/Standards/E2473.htm.
3. ASTM International. ASTM E2084-00, Standard Specification for Authentication of Healthcare Information Using Digital Signatures. Available at https://www.astm.org/DATABASE.CART/WITHDRAWN/E2084.htm.

[a] Kasthurirathne, S. N., B. Mamlin, G. Grieve, and P. Biondich. "Towards Standardized Patient Data Exchange: Integrating a FHIR Based API for the open medical record system. *Studies in Health Technology and Informatics* 216 (2015): 932.
[b] Oh, A. S. "A Study on HL7 Standard Message for Healthcare System Based on ISO/IEEE 11073." *International Journal of Smart Home* 9, no. 6 (2015): 113–18.
[c] Meehan, R. A., D. T. Mon, K. M. Kelly, M. Rocca, G. Dickinson, J. Ritter, and C. M. Johnson. "Increasing EHR System Usability through Standards: Conformance Criteria in the HL7 EHR-system Functional Model." *Journal of Biomedical Informatics* 63 (2016): 169–73.

Table 3

Format of the Main Data Set Categories Extracted from the Standards and Paper-based Record Sheets

Data Class	ASTM Standards		HL7 Standards		Paper Record	
	Free text	**Structured**	**Free text**	**Structured**	**Free Text**	**Structured**
Administrative	14	9	17	6	9	3
Encounter	6	4	-	3	6	6
Problem	4	2	-	-	-	-
Treatment plan	7	-	3	-	1	-
Provider	1	-	1	-	3	-
Exam	3	1	1	-	5	3
Diagnosis	2	9	-	-	3	-
History	3	4	-	-	6	-
Event	3	1	2	1	2	-
Insurance	2	3	5	3	6	1
Total	45	33	29	13	41	13