Data incompleteness preventing information communication from hospital information systems to the Iranian national electronic health record (SEPAS)

Reza Abbasi1, Reza Khajouei2*, Monireh Sadeghi Jabali3, Moghadameh Mirzaei1

1Department of Health Information Management & Technology, Faculty of Paramedicine, Kashan University of Medical Sciences, Kashan, Iran
2Department of Health Information Sciences, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran
3Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

**A B S T R A C T**

**Introduction:** One of the well-known problems related to the information quality is the information incompleteness in health information systems. The purpose of this study was to investigate the completeness rate of patients' information recorded in the hospital information system, sending information from which to Iranian electronic health record system (SEPAS) seemed to be unsuccessful.

**Material and Methods:** This study was conducted in six hospitals affiliated with Kerman University of Medical Sciences (KUMS) in Iran. In this study, 882 records which had failed to be sent from three hospital information systems to SEPAS were reviewed and data were collected using a checklist. Data were analyzed using descriptive and inferential statistics with SPSS 18.

**Results:** A total of 18758 demographic and clinical information elements were examined. The rate of completeness was 55%. The highest completeness rate of demographic information was related to name, surname, gender, nationality, date of birth, father's name, marital status, place of residence, telephone number (79-100%), and in clinical information it was related to the final diagnosis (74%). The completeness rate of some information elements was significantly different among the hospitals (p <0.05). The completeness rate of information communicated to the Iranian national electronic health record was at a moderate level.

**Conclusion:** This study showed that completeness rate is different among hospitals using the same hospital information system. The results of this study can help the health policymakers and developers of the national electronic health record in developing countries to improve completeness rate and also information quality in health information systems.

Cite this paper as:

**INTRODUCTION**

Almost since 2005, most health care institutions have been implementing a variety of electronic documentation systems and electronic health records [1]. Most health policy makers emphasize the importance of electronic health records, exchange and proper use of information and types of communication technologies in improving the coordination, quality and efficiency of the health care processes [2-4]. Collecting the clinical information from the electronic health record and maintaining them in larger databases such as databases in the ministry of health (MOH) and using them to improve clinical decision-making and care quality in health centers as well as reducing errors in drug prescription and conducting research are on the rise [5, 6]. Developing countries such as Iran, after computerizing process of health care system, have implemented the electronic health record system and all community individuals have used smart electronic health cards. This system is called "Iranian electronic health record" (SEPAS) [7, 8].
In recent years, despite the adoption and implementation of electronic health record (EHR) in health centers, the predicted goals for health care have not been completely met. One of the main reasons for this failure was insufficient quality of the main part of the information collected by EHR [5, 9-11].

Information quality refers to the use of complete, accurate, understandable, and accessible information when it is needed and its presentation in a format that the user can easily use as well as making sure that information is collected regarding the purpose [12]. Clinical information systems require complete and accurate data to provide the best possible services [13]. Poor data quality can be a serious threat to the validity and generalizability of the clinical research findings [14]. Therefore, given the importance of information in the electronic health record in the provision of health care services, this information should be of sufficient quality and quantity [15]. Promoting and developing information quality contained in the electronic health record will help increase confidence in the use of this information [13]. It is anticipated that lack of quality and quantity of information in EHR after its implementation has led us to misunderstand the significant changes emerging from using EHRs in improving the quality of health care services [5, 10, 11, 16-18]. For this reason, it is very important to monitor and improve information quality and quantity in order to understand the condition of health information collected in the EHR [6]. Information completeness rate is known as one of the key indicators of information quality [2].

Incompleteness of data elements (DEs) in the health information systems is a major and recognized problem affecting the quality of health care provision [10, 19-21]. Improving information completeness in the patients' electronic records is one of the steps in improving information quality and providing health services [21].

Most studies have examined the impact of electronic records on the rate of information completeness [7, 19, 21-23], and few other studies have looked at the extent of completeness of each information element in EHRs [24]. Iran is also a developing country, and so far, there has been no study on the extent to which information in the health information systems has been completed in this country.

In this country, data elements in patients' electronic records are sent from hospital information systems to the Iranian national electronic health record. According to a study [25], one of the factors possibly leading to failure to communicate information among health information systems is incompleteness of information elements. Therefore, the purpose of the present study was to investigate information completeness rate unsent from hospital information systems to the Iranian national electronic health record.

**MATERIAL AND METHODS**

**Study design and setting**

This analytical study was conducted in six hospitals affiliated with Kerman University of Medical Sciences (KUMS) in Iran. In this study, the information of patients' records failed to be communicated from three common hospital information systems to Iranian national electronic health record were investigated.

**Sample size determination**

In total, 6334 records had not been communicated to the SEPAS during this period. The sample size was estimated as 882 records selected from each hospital based on the proportional allocation. The included records were randomly chosen from the list of the not communicated records in each hospital, with an interval of 7 records.

**Data collection**

According to the guidelines issued by the MOH in Iran [7] on the DEs communicated from hospital information systems to the electronic health record, a checklist was made to determine the completeness and incompleteness of 33 DEs (22 and 11 demographic and clinical DEs) sent from these systems to the EHR. Demographic data elements consisted of 14 individual DEs and 8 DEs related to the occupation and place of residence. Data elements submitted to the Iranian national electronic health record included two categories of mandatory and optional DEs. It is necessary to send mandatory data elements to the EHR, while it is unnecessary to send optional data elements. The validity of this checklist was reviewed and approved by three medical informatics specialists. After the initial review of the hospital information systems investigated, it was determined that there were 8 optional DEs out of the 33 DEs in the ministry of health's instructions, such as father's surname, patient's job description, diagnosis date, diagnosis time, diagnosis of severity, diagnosis condition, location of patient's death, and the ward in which the patient has died, which did not exist as the minimum data in the examined systems. These DEs were excluded from the study and their completeness rate was not calculated. Therefore, in this study, the completeness rate of 25 DEs was studied.

**Data analysis**

Data were analyzed using SPSS (ver. 18) through descriptive and analytical statistics. Analytical statistics such as Chi-square and Fisher were used to
investigate the relationship between completeness and incompleteness of the examined DEs in the electronic health record with the type of hospital.

**RESULTS**

In general, completeness rate of the 20 demographic data elements and 5 clinical data elements sent to the health electronic record was estimated at 55%. Table 1 shows the rate of completeness and incompleteness of the demographic and clinical data elements separately.

**Table 1: The completeness rate of the demographic and clinical data elements**

<table>
<thead>
<tr>
<th>Data elements</th>
<th>Completeness rate (%)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic data</strong></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>99</td>
</tr>
<tr>
<td>Sure name</td>
<td>100</td>
</tr>
<tr>
<td>Father’s name</td>
<td>98.5</td>
</tr>
<tr>
<td>Mother’s name</td>
<td>0.5</td>
</tr>
<tr>
<td>Mother’s sure name</td>
<td>2</td>
</tr>
<tr>
<td>National code (M)</td>
<td>80</td>
</tr>
<tr>
<td>ID number</td>
<td>50</td>
</tr>
<tr>
<td>ID number place</td>
<td>13</td>
</tr>
<tr>
<td>Birthday</td>
<td>99</td>
</tr>
<tr>
<td>Gender</td>
<td>97</td>
</tr>
<tr>
<td>Marital status</td>
<td>73</td>
</tr>
<tr>
<td>Nationality</td>
<td>96.5</td>
</tr>
<tr>
<td>Passport number</td>
<td>0.1</td>
</tr>
<tr>
<td>Place of residence</td>
<td>61</td>
</tr>
<tr>
<td>Full address</td>
<td>92.5</td>
</tr>
<tr>
<td>Postal code</td>
<td>79</td>
</tr>
<tr>
<td>Home telephone number</td>
<td>21</td>
</tr>
<tr>
<td>Mobile number</td>
<td>84</td>
</tr>
<tr>
<td>Education level</td>
<td>98</td>
</tr>
<tr>
<td>Job</td>
<td>6</td>
</tr>
<tr>
<td><strong>Clinical data</strong></td>
<td></td>
</tr>
<tr>
<td>Final diagnosis (M)</td>
<td>74</td>
</tr>
<tr>
<td>Date of death</td>
<td>85</td>
</tr>
<tr>
<td>Death time</td>
<td>12</td>
</tr>
<tr>
<td>Cause of death (M)</td>
<td>58</td>
</tr>
<tr>
<td>Underlying cause of death</td>
<td>27</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>55</td>
</tr>
</tbody>
</table>

* The rest were incomplete  M: Mandatory data element

Accordingly, in general, in these hospitals, the rate of completeness and transmission of the identification data elements such as name, first name, family name, father’s name, gender, date of birth and patient nationality were in the range of (96.5–100%) and data elements such as national code and marital status were about (70–80%).

Furthermore, the results showed that, except for the data elements related to the birth date and passport number, the rate of completeness and incompleteness of other data elements in the hospitals and the studied information systems were significantly different (p<0.046).

In Table 1, completeness and incompleteness rate of the data regarding occupation and address are shown. Accordingly, the highest data completeness rate relates to patient’s place of residence, full address and telephone number and is within the range of 85-92%. According to the results of this study, except for the postal code, the completeness and incompleteness rate of the other data elements in the hospitals and the information systems were significantly different (P<0.02).

The highest completeness rate of this data was related to the final diagnosis, date of death and the underlying cause of death, being in the range of 73-85%.

There was also a significant difference between the completeness and incompleteness rate of the final diagnosis, date, time, cause and underlying cause of death in studied hospitals and the information systems (P<0.045).

According to the findings of this study, about 253 data elements were filled with ‘0’ (Fig 1). In a fraction of 16% of cases, ID number element was filled with ‘0’ and in about 1.5-5.4% of cases, national code, home phone number, level of education and patient occupation were filled with ‘0’. Also, as shown in Fig 2, some data elements were written with misspellings (n=13) or less or more characters (n=32) by users.

**DISCUSSION**

The results of this study showed that the minimum data set in the three examined hospital information systems was different. For example, the data elements of the patient’s mother’s first name and surname, occupation, level of education, place of birth, place of residence, postcode and time of death were not only defined identically and equally in all three systems, but also in some of them, the minimum data did not even exist.

**Comparisons with existing literature**

In this study, the total of completeness rate of demographic information was 52% and less than the completeness rate of demographic information in electronic patients’ record in Ethiopia [26]. Since it is necessary to record demographic information for clinical procedures and meaningful use of electronic health records [27], it is important for users to pay attention more to the completion of this information. The highest completeness rate of the demographic data elements was related to patients’ name, surname, gender, nationality, date of birth, father’s name, place of residence, mobile number, marital status and national ID number. In line with our study...
findings, the results of studies which was investigated the completeness rate in electronic health records showed some data elements such as gender, age, nationality, marital status of people had a high completeness rate [6, 24, 26, 28, 29]. It seems to be common of this information is one of the probable reasons for high completeness rate of them. Patients can provide this information to health information system users easily and without identification card. Adding the interoperability of information capability with other databases such as registry office and insurance organizations and mandatory recording of information by users may be effective in improving the completeness of the information.

Fig 1: Data entry fields filled with '0' (in Persian)

Fig 2: Inaccuracy of data entry due to spelling mistakes or fewer and more characters (in Persian): B) Fewer or more characters than usual. C) Failure to accurately record information in the information fields. D) Spelling mistakes while typing the information

The results of this study showed the highest completeness rate of the clinical data elements was related to final diagnosis, date and cause of death and the lowest was related to time and underlying cause of disease. Parkin [30] studied data completeness rate in cancer record systems in different countries such as Germany, France, England, Japan and Korea. Data completeness rate on patients’ death in these systems was in the range of about 50-85%. Since clinical data collecting is a part of patients’ care process, this data has a relatively high completeness rate usually [31].

The results of this study showed that about a quarter of the data elements requested by the patient such as father's surname, job description, diagnosis date, diagnosis time, diagnosis severity, diagnosis status, and place of death were not reported in any of the three HISs and some of the other data elements such as postcode, occupation, level of education, passport number, place of birth, mother's name, and surname were not available. Currently, it is not mandatory to record these data elements in SEPAS. Therefore, it may be possible to increase data completeness by standardizing the minimum data set in HISs and also by requiring the registration of data elements [32].

Based on the results of the present study, the presence of spelling mistakes, fewer and more characters, or filling the fields with '0' are the most important instances of the inaccurate data. One of the possible reasons for this problem can be because of the inaccuracy of the users in data recording. Also, in order to avoid some of the system messages containing the data elements, fields are left blank. Some users fill out some of them with unrelated characters such as '0' or some unreasonable characters in order not to encounter any problems in the process of entering and recording the data. Some studies [6, 33] revealed that training the users simultaneously or periodically in reviewing data quality has a significant impact on data completeness rate and accuracy. Also, the results of Verkerke's study on Nursing Information System [34] and Majeed's on EHRs [35] suggested that the accurate and complete recording of data depends largely on the users of these systems and their passion for recording the data.

Therefore, in addition to training users, in order to improve the quality and quantity of the data documentation, it may be better to encourage them to record the data accurately and completely and help them properly understand its importance.

Another reason for the inaccurately recording and filling up the data fields with '0' or other unreasonable characters by the users may be a lot of workload, followed by users’ rush to record the data. According to the results of Forster's study [36], the time consumed by the users of HISs to enter and collect data is correlated with the data completeness. Increasing the number of staff at different hospital units responsible for recording patients’ data in HISs may also reduce the workload to a certain extent.

In addition, designing systems and applying systems with high usability may also affect data quantity and quality. Arts believed the poor designing of systems lead to incompleteness of data [37]. It would be better to design the systems in such a way as to avoid entering inaccurate and unreasonable data and warn the users not to enter such data or prevent them from going to the next stage [38, 39]. Byrd also pointed out in his study [15] that the reason for increased data quality in the new HIS compared to the old one is considering factors such as design and use of structured data elements and sliding menus as well as recording the data at the moment and on patients’ bedside by service providers. Wells [40] also
recommended recording completely structured data and using sliding menus to reduce the missing data.

The results of this study showed that the completeness rate of data elements in HISs as well as the variety of the investigated hospitals are different. The results of two systematic studies \[11, 41\] also indicated that rate of completeness in HISs varies. Factors such as differences in usability of systems, lack of a unified definition, using the same minimum data set as well as the difference between users of these systems may affect information completeness rate.

In the present study, a large number of the fields related to the demographic data, occupation, home phone number, and place of residence were incomplete. As with the information systems of the Medicaid and Medicare institutions that were linked to other information systems \[2\] to access and collect the patients' information, information system designers can also automatically collect these data elements by coordinating and connecting their HISs with information systems of other centers such as the registry office, labor offices, insurance organizations, forensic medicine organization as well as the statistics office. Also, given that the Unified Health Identifier (UHID), with the use of community biometrics, helps to improve patient safety and information accuracy to be used in HISs such as EHR \[42\], it may be possible to use this technology to improve the quality of people's identification information.

The high quality of data is prior to its use \[15\]. Analysis of the existing data in HISs is important in reporting, modeling and predicting outcomes, monitoring public health, improving the quality of health care services, conducting clinical research, decision-making purposes, policy-making and health system management \[6, 43\]. Since the EHR is recognized as a key tool in improving the quality of health care and is used to access complete, accurate and reliable data, health care providers should tackle it with more care and higher priority.

Iran is a developing country where designing and application of HISs as well as investigating and improving the data quality in these systems have a short history of about a decade. Therefore, newness of designing industry, implementation, and use of HISs in Iran and other developing countries can be regarded as one of the reasons for weakness in quantity and quality of the data recorded in these systems. In many countries with the EHRs, there is a lot of information exchange between HISs and national health record on a daily basis. In Iran, which is known as a developing country, there are also numerous HISs, from which a lot of information is daily transmitted to the national health record of the country. Although more precise results are expected to be obtained through this study at a wider level, the current study is conducted on the hospitals in one of the metropolises, as well as the most widely used hospital information systems in a developing country. Therefore, the results of the present study could be generalized to the HISs in other developing and developed countries as well.

Moreover, the results of the present study showed that there is a significant difference between completeness and incompleteness rate of some data elements in investigated hospitals, some of which used the same HIS. This correlation seems to be due to the fact that most investigated hospitals are considered public hospitals and patients with different nationalities or various diseases refer to these hospitals and receive several health services. There may be a correlation between data completeness rate and some other data elements such as nationality or the type of disease. Thus, it is suggested that this topic be addressed in future studies.

The study found that more than half of data elements sent from HISs to the national electronic health record were complete. Based on the results of this study, hospitals that use the same HIS also vary in their rate of completeness. Therefore, it may be concluded that data completeness rate relates only to the design of information systems and the efforts and accuracy of users when entering data is also very effective. In addition to users' attention while entering data, it is better to design the systems in such a way as to facilitate the entry of the information and prevent entering incorrect data.

**Limitations and future recommendations**

There is one major limitation in the current study. There are several HISs used in different Iran's hospitals sending a lot of hospital records to the Iranian national electronic health record every day. Nevertheless, we investigated three most commonly used hospital information systems among the 6 major hospitals of KUMS. Future investigation on other hospital information systems may lead to more generalizable and precise findings.

**Implications for research and practice**

The results of this study can be used in developing countries for optimizing the quality and quantity of the health information systems. Also, these results draw the attention of the designers and developers of the health information systems to the situations resulting in incompleteness and inaccuracy of data and help them to prevent these situations in the design phase of the systems. It can also help the hospital managers and the authorities to monitor and improve the quality and quantity of data and system reports. Completeness of data at the point of first data entry can improve the quality of data recorded in information systems and ultimately communicated.
to the larger data storages, such as the ministry of health data warehouses. This type of data can be used as a reliable source for policy-making, planning and research in the health-related fields.

CONCLUSION

The results of this study showed that the data completeness rate in Iran’s national EHR was moderate. It was also found that the type of hospitals, system users, and various aspects of hospital information systems such as their minimum data set or the selective ness of some data entry fields can affect the completeness rate of data.

Since the completion rate of data was different in the hospitals using the same HIS, it can be concluded that completeness rate does not necessarily depend on the design of the information systems. This highlights the significant role of the users in recording the data. Therefore, based on these results, we recommend the following two categories of solutions for improving data completeness rate. The first category is human related solutions such as increasing human resources to reduce the work pressure on users, educating and encouraging them to record complete, correct and timely information in the system, conducting periodic quality reviews of EHRs and determining a specific time frame for recording information in systems in a timely manner. The second category is the systematic modifications of systems and settings such as using the same minimum data set for all the health information systems, mandatory recording of data in the system, providing system feedbacks and error messages, using sliding menus in order to reduce data entry by users, encouraging structured data entry, connecting information systems to the database of some organizations such as registration office, forensic organization, insurance centers and other health care centers to obtain identification, clinical, and death information of patients, and using biometric technologies to identify the patients. More studies are required to evaluate the true effect of these modifications on the accuracy and completeness of data communicated among different systems.

AUTHOR’S CONTRIBUTION

RA, RK, and MSJ contributed to the conception and design of the study, acquisition and interpretation of the data, and drafting the paper. MM and RA were primarily responsible for the statistical analysis of the data.

All authors contributed to the literature review, design, data collection and analysis, drafting the manuscript, read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

ETHICS STATEMENT

This work was approved by ethics committee of Kerman University of Medical Sciences (IR.kmu.REC.1394.333)

REFERENCES


32. Abbasi R, Khajouei R, Mirzaee M. Evaluating the


