Establishing core data elements for colorectal cancer and mapping to FHIR resources: Towards interoperable Iranian health information systems

Raoof Nopour, Somayeh Nasiri, Maryam Ahmadi, Mohammad Shirkhoda[1]

Abstract:

BACKGROUND: Colorectal cancer (CRC) has a high prevalence, threatening public health. Despite the high prevalence, there has not been a proper solution to standardize the data in this domain in Iran. Therefore, this study aims to provide a new approach to standardizing CRC data using Fast Healthcare Interoperability Resources (FHIR).

MATERIAL AND METHODS: This study was conducted in five phases. First, a literature review was performed to gain the CRC core data elements based on the scientific databases and thematic websites. Second, we developed a researcher-made questionnaire with a five-choice Likert scale to gain the most important core data elements based on the opinions of specialists. Third, we leveraged the Delphi method to score these elements and develop the minimum data set (MDS) for CRC. Fourth, we used the terminologies to better perceive the concepts and thematic classification of core data elements. Fifth, we leveraged the FHIR to map the CRC core data elements to resources and standardize them.

RESULTS: Ten CRC data categories of classes were obtained from the literature review and Delphi, including administrative, diagnosis, therapeutic procedure, follow-up, patient referral, vital status, drug, clinical assessment, signs and symptoms, and history of conditions. The obtained resources were patient, care team, condition, family member history, medication request, adverse event, schedule, procedure, location, observation, diagnostic report, specimen, molecular sequence, genomic study, and service request.

CONCLUSION: This study showed the FHIR-based standardization of the CRC data as an approach for integrating health institutions to facilitate clinical decision-making and secondary use of healthcare data.

Keywords:

Colorectal neoplasms, common data elements, health information interoperability, reference standards, systems integration

Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran, ¹Cancer Research Center, Cancer Institute of Iran, Tehran University of Medical Sciences, Tehran, Iran

Address for Correspondence:
Dr. Maryam Ahmadi, Ph.D. of Health Information Management, Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran. E-mail: ahmadi.m@iums. ac.ir

Introduction

Colorectal cancer (CRC) is the growth of cancerous masses in the colon and rectum tissues in the bowel, which can spread to other body tissues.[1] This cancer group is the third most common malignancy and the fourth most cancer-associated death most common, endangering public health conditions globally.[2,3] CRC accounts for more than 1,850,000 new cases and an 850,000 mortality rate.[4] The highest CRC occurrences were reported in countries with favorable economic conditions.[5] Although modifiable and non-modifiable factors usually determine CRC, more than half of the morbidity and mortality associated with CRC are adjustable.[6,7] Modifiable

categories include obesity, smoking, physical activity, alcohol consumption, and nourishment, which are more prevalent in developed countries.[8,9]

It is estimated that the overall percentage of CRC-associated deaths will increase to 60% and 71% related to rectum and colon cancers by 2035.[10] In the United States, approximately 147,950 people were detected as new cases, and 53,200 patients died from CRC, the third deadliest cancer among men and women, in 2020.[11] Also, the high CRC incidence rate was reported in other developed countries such as Europe and Japan.[12]

In the recent decade, developing countries have experienced an increasing trends of CRC due to the Westernized lifestyle.[13] In Iran, CRC has been seen as an upward trend in the mortality and morbidity rate in recent years, and this is the third and fourth most prevalent cancer among men and women, respectively.[14]

The survival rate of CRC in the early stages is considerably higher than advanced ones, so leveraging early and suitable screening and management approaches is crucial.[15] Also, considering the increasing trend of death-related CRC, using appropriate strategies significantly impacts the efficient controlling and management of this disease and increases the CRC-related survival rate.[16,17] Previous studies have shown that informatics solutions could be leveraged as crucial tools for screening and managing CRC.[18-20]

Using various health information systems (HISs) significantly enhances the quality of care in healthcare systems.[21] These systems also establish surveillance ability in healthcare institutions and research environments to manage CRC.[22]

Designing a database is one of the essential steps in implementing HISs.[23] A well-designed database with standardized data elements can facilitate uniform data collection in HISs by establishing a minimum data set (MDS).[24,25] However, we need more technical requirements to achieve interoperability among various HISs, especially electronic health records (EHRs).[26]

By increasing the use of various digital health interventions in sophisticated healthcare settings, a lack of healthcare data standardization incurs that healthcare providers cannot share healthcare data between various institutions.[27] Also, without any data standardization, the sharing of health information is hampered, and the reusability of data is decreased, which causes the existence of scattered databases without any data exchangeability.[28] It also causes different HISs associated with healthcare organizations to be separated,

decreasing the quality of care and surveillance targets that impede the adoption of E-health targets in Iran.[29]

As a standard, the FHIRs provide interoperability by facilitating clinical data sharing between HISs through mapping clinical data elements to information units, namely resources.[30,31] Some advantages of FHIR consist of being the restful configuration paradigm, the easiness of medical concept perception, the usability of mobile tools, and facilitating the data exchange between different HISs in clinical and research settings.[32]

So far, several studies have been conducted in Iran to integrate various cancer databases by establishing uniform core data elements. They mentioned that developing a core primary data set has a significant role in uniform collecting, reporting, and exchanging data between various HISs.[33-35] Developing MDS is the fundamental level of interoperability between various HISs,[36] and we need more powerful interoperability solutions in this regard. Also, some works have been conducted on providing interoperability solutions to integrate various HISs based on the (clinical document architecture) CDA as a document-centric standard.[37-39] This standard is more common in use when dealing with more complex structures of clinical data in extensible markup language (XML) format.[40] To use the clinical data at an individual level, we require the standards with a more granular structure to share the health data between various HISs to EHRs.[41] The FHIR standard, by leveraging the information representation approach as resources, has the capability of data exchange with high granularity between various HISs.[42] Considering an increasing in the volume of healthcare data and various HISs, especially the EHR in healthcare environments, the efficient data exchange to these HISs has a crucial role in promoting data interoperability and quality of care in various clinical care settings.[43] The FHIR-based data standardization is advantageous for exchanging healthcare data, especially for EHRs.[31] Based on the previous studies conducted, there has been no topic on leveraging the FHIR standard in the cancer domain in Iran, and determining the uniform data elements in this way is the first time being performed in the cancer domain in this country. So, the novelty of this study is to build the standardized data elements for CRC to achieve enhanced interoperability purposes, especially to facilitate data exchangeability to the EHR system. As a pioneering study in the cancer field in Iran using FHIR, this research aims to achieve a more comprehensive approach for data standardization in the cancer domain, especially CRC. To this aim, we first establish a core data element for CRC and then conceptualize and map them to FHIR resources to make CRC standardized data elements having more data exchangeability to various HISs, especially EHRs.

Material and Methods

Overview of the study

This study was an applied approach that was performed in five steps as shown in Figure 1.

Literature review

In this phase of the study, we searched the scientific databases to extract relevant CRC core data elements until data saturation. To this end, we investigated the Web of Science (WOS), Scopus, PubMed, Science Direct, Magiran, Google scholar, Islamic World Science Citation Center (ISC), Proquest, Scientific Information Database (SID) in advanced search mode, and thematic websites through a mixture of keywords, including "minimum data set," "minimum data element," "core data set," "core data element," "essential data set," "essential data element," "common data element," "data set," "registry system," "database system", "health information system," "colorectal cancer," " colorectal neoplasm" and "hospital information system" and Boolean logic of "and," "or," and "NOT." The truncation operator, quotation, and limitation in combination with Mesh terms were used for searching purposes. The publications written in English and Persian languages, having a content relationship with the research domain, and were published on dates ranging from 2010 to 2023, were included in this study.

Questionnaire development

After extracting CRC core data elements from scientific databases and thematic websites, we developed a researcher-made questionnaire to rank CRC core data elements based on the panel of experts. Each extracted CRC core data element was entered into the questionnaire. We listed all CRC core data elements in one column and assigned five columns to five Likert spectrums to measure the degree of importance of each CRC core data element. Scores from one to five were assigned to each component's lowest to highest degree of importance. The content validity of the questionnaire was evaluated by five experts, including two gastrointestinal (GI) and three oncology specialists.

Delphi phase

We leveraged the Delphi method to obtain the opinions of the panel of experts on CRC core data elements and finalize the MDS for CRC. The panel of experts included 30 specialists in GI and oncology. Clinically, the polling process was performed by specialists who had at least five years of experience working in clinical settings and were familiar with medical records and CRC core data elements used in HISs. So, the core data elements selection was based on the clinical medical records systems and HISs, in addition to the opinions of experts. The demographic and occupational characteristics of the panel of experts in the first and second rounds of the Delphi technique are shown in Table 1. To this end, we used the two-round Delphi method and finalized the CRC MDS. The core data elements in the first round of the Delphi method that obtained 75% and higher of the panel of experts' agreement, were leveraged as CRC minimum data elements in the study. On the contrary, the core data elements with less than 50% of the accords were excluded from the study in this round. Since some core data elements had 50%–75% specialists' contraction, we used the second Delphi round for those elements, and the elements with higher 75% agreement in the second round were considered for CRC MDS.

Finalization and perception of minimum data elements

In this study, the final MDS for CRC was obtained as the results of the first and second phases of the Delphi method. Each core data element was assigned to a category and subcategory of class based on the conventional qualitative content analysis method and clinical experts' opinions. To better understand the concept of each CRC core data element, We investigated articles on CRC core data elements, CRC diagnostic, therapeutic, and screening guidelines, thematic websites, Unified Medical Language Systems (UMLSs), and searched CRC core data elements in the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) browser [Figure 2].[44] Also, we found the precise concept of each CRC core data element by searching and investigating the parent and child levels of it in

Figure 1: The methodology map of the study

Table 1: Some characteristics of the expert panel

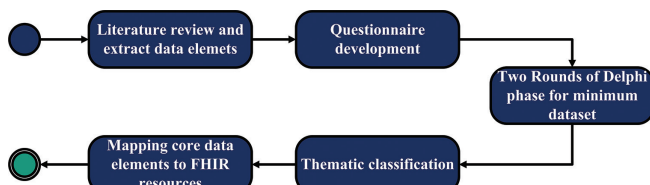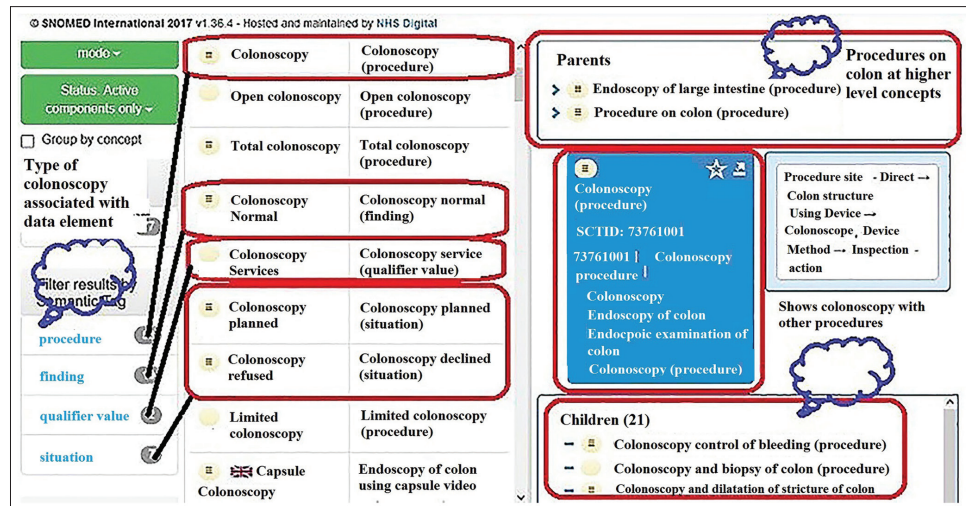| Characteristics		Number of participants
		First and second rounds
Specialty	Gastrointestinal	12 (0.4)
	Oncology	18 (0.6)
Gender	Male	12 (0.4)
	Female	18 (0.6)
Age group (years)	30-40	3 (0.1)
	40-50	8 (0.27)
	50-60	10 (0.33)
	>60	9 (0.3)
Work experience (years)	<10	4 (0.13)
	10-20	7 (0.23)
	20-30	10 (0.33)
	>30	9 (0.3)

Figure 2: A schema of SNOMED-CT browser associated with colonoscopy

SNOMED-CT. We performed this process to better assign relevant resources in FHIR considering the hierarchical concepts in SNOMED-CT, consequently leading to better-assigning resources through better concept perception.

Mapping CRC core data elements to FHIR resources

We mapped each CRC core data element to the more relevant FHIR resources. The mapping process was performed in each category or subcategory of CRC data one by one, and each core data element that existed in the each category or subcategory of class, was placed in the relevant resource. The mapping process was done by one of the authors using the FHIR website,[45] and another author checked the accuracy of the mapping process from each CRC core data element to the associated resource.

Results

Identifying CRC core data elements

By searching CRC core data elements in scientific databases and thematic websites, we obtained 189 core data elements. The CRC core data elements were categorized into two classes, including non-clinical and clinical. CRC core data elements were classified into 10 categories of classes: diagnosis, therapeutic procedure, follow-up, patient referral, clinical assessment, history of conditions, signs and symptoms, administrative, drug, and patient status. The results of two rounds of the Delphi method for CRC core data elements extracted from the literature review are shown in Table 2 according to the data classes' categorization.

Based on the information in Table 2, from the total of 189 CRC core data elements obtained from a literature review, 22 were removed through the opinions of the panel of experts. After the first round of the Delphi phase, 11 core data elements were recommended by specialists and were entered into the second round. They included name of provider (administrative category), date of diagnosis, date of sampling and imaging (diagnosis), date of surgery and name of surgeon (therapeutic procedure), date of follow-up (follow-up), date of therapeutic plan (patient referral), date of examination (clinical assessment), family history of polyp (history of conditions), and feeling of not emptying bowels (signs and symptoms).

Core data elements were released in the first (n = 13) and second (n = 9) rounds of Delphi. Forty-six core data elements of the diagnosis category remained in the study. Thirteen core data elements of therapeutic procedures remained in the study. Follow-up and patient referrals included 17 and seven core data elements, respectively. For clinical assessment, signs and symptoms, and history of conditions, we acquired eight, 14, and 41 core data elements, respectively. Twenty core data elements were considered for administrative data. Drug data and patient status without any core data element deletion obtained eight and four core data elements, respectively. Overall, we gained 178 CRC core data elements in 10 categories of classes, and they were considered for the final CRC MDS. We presented some of the core data elements belonging to 10 categories of classes in Table 3.

Table 3 shows the final MDS for CRC. The MDS was categorized into 10 and 27 categories and subcategories of classes, respectively. The diagnosis category includes data on status, diagnostic methods, diagnostic techniques, imaging, laboratory tests, tumor markers, molecular and genomic profiles, and pathological findings. Therapeutic approaches also in the CRC MDS consisted of two subcategories of radiotherapy and surgery. The follow-up included three subcategories: general contact, recurrence and plan. The patient referral included preoperative and

Table 2: The results of the two-round Delphi expert panel

Data categories	Number of data elements	The first round of Delphi			The second round of Delphi			The final number of data elements
		<50%	50%-75%	>75%	<50%	50%-75%	>75%	
Diagnosis	47	3	1	43	1	0	3	46
Therapeutic procedure	13	2	0	11	0	0	2	13
Follow-up	17	0	2	15	1	0	2	17
Patient referral	7	1	0	6	0	0	1	7
Clinical assessment	11	2	2	7	2	0	1	8
History of conditions	45	3	2	40	2	0	1	41
Signs and symptoms	15	1	1	13	1	0	1	14
Administrative	22	1	2	19	2	0	1	20
Drug	8	0	0	8	0	0	0	8
Patient status	4	0	0	4	0	0	0	4
Total	189	13	13	166	9	0	12	178

Table 3: The final minimum data set for CRC

Category	Subcategory	Data elements
Diagnosis	Date of status	Date of diagnosis, date of first metastasis, days from other events to diagnosis
	Diagnostic methods	Basis of the diagnosis of tumor, route to diagnosis
	Diagnostic techniques	Clinical investigation, laboratory test, image investigation
	Imaging	Imaging technique, date of imaging, diagnostic results
	Laboratory test	Date of sampling, type of test, test results
	Tumor markers	Marker names, method of analysis
	Molecular, genomic profile	Molecular diagnostic findings, molecular pathologic findings
	Pathological findings	Tumor site, specimen integrity, polyp size
Therapeutic procedure	Radiotherapy	Date of starting procedure, date of ending practice, dose of radiotherapy, post-treatment complication
	Surgery	Surgery type, surgery method, surgery site, date of surgery, surgeon name, surgeon ID, clinical setting type
Follow-up	General contact	Date of last contact/death, date of follow-up, vital status
	Recurrence	Date of first recurrence, type of first recurrence, recurrence site, distance recurrence site
	Plan	Subsequent treatment, date of treatment
Referral	Preoperative	Cause of referral, outpatient clinical investigation, events, date of the plan
	Postoperative	Outcomes, feedback to General Practitioner (GP), measures after discharge
Clinical assessment		Abdomen mass, weight, American society of anesthesiologists (ASA) score, body temperature
History of conditions	Personal history	Inflammatory bowel disease, surgery of gastrointestinal organs, diabetes,
	Family history of other diseases	Polyp, cystic fibrosis, heart disease,
	Family history of cancer	Breast, thyroid, cervix, uterus, digestive system
Signs and symptoms	General	Abdominal pain, abdominal sensitivity, loss of weight, fatigue
	Gastrointestinal	Intestinal obstruction, intestinal perforation
Administrative	Patient data	National code, name, last name, father's name
	Care provider institute data	Care provider name, care provider address, phone number, home address
Drug	Immunotherapy	Drugs name, adverse effect
	Chemotherapy	Preoperative drugs name, postoperative drugs name
	Current drug	Drugs name, adverse effect
Patient status		Vital status, cause of death, Carnovsky performance situation

postoperative situations subcategories. History of conditions in the CRC MDS included three subcategories: personal history, family history of other diseases, and family history of cancer. The signs and symptoms were categorized into two subcategories: gastrointestinal and general signs and symptoms. Administrative data included two subcategories of patient and care provider institute data. Also, drug information was classified into three immunotherapy, chemotherapy, and current drug subcategories.

Mapping CRC core data elements to FHIR resources

The mapping process in this study was depended on two conditions: using different resources for different states of core data elements; for example, for each imaging technique, we used various ("procedure") resources, or for each sign and symptom, such as fatigue and abdominal pain. Some core data elements

had different natures in one category or subcategory; for example, some elements defined the specimen characteristics in the pathology data category, and others were described in the diagnostic results. In this situation, we used two different resources ("specimen") and ("diagnostic report") for defining these core data elements, respectively. The results of mapping some CRC core data elements to FHIR resources are represented in Table 4. In this regard, we presented some CRC core data elements in each category of class.

Based on Table 4, we observed that most data types are codable concept; for example, we could point to the history of conditions and sign and symptoms categories. The codable concept type of each resource data element based on the SNOMED-CT codes could support semantic interoperability in addition to structural type, and we concluded that based on the mapping results, many core data elements are codable. For some core data elements, such as the outcome of surgery in the patient referral category, we could not find any relevant data elements in the service request, so we added this item as an extension to this resource.

Generally, based on the study, we obtained 15 FHIR resources, including patient, care team, condition, family member history, medication request, adverse event, schedule, procedure, location, observation, diagnostic report, specimen, molecular sequence, genomic study, and service request as the results of CRC data mapping. The general schema of mapping CRC core data elements is shown in Figure 3.

Discussion

This study aims to introduce a solution for efficient CRC data exchangeability between various HISs in the cancer domain. In this regard, we first developed an MDS for CRC using a literature review and the Delphi technique. In the next step, we mapped all extracted CRC core data elements to the relevant resources in establishing the standardized CRC core data elements according to the FHIR standard and enhancing the interoperability of various HISs, especially EHRs. Based on the literature review and utilizing the experts' opinions, we identified 10 CRC categories of classes, including administrative, diagnosis, therapeutic procedure, follow-up, patient referral, clinical assessment, history of condition, signs and symptoms, drug, and patient status. As the results of mapping the CRC core data elements, 15 FHIR were obtained, including patient, care team, condition, family member history, medication request, adverse event, schedule, procedure, location, observation, diagnostic report, specimen, molecular sequence, genomic study, and service request.

The data exchangeability between various HISs having standardized CRC data mapped to FHIR resources, allows communication between multiple healthcare ISs efficiently.[27] With these mentioned capabilities in already isolated healthcare institutions, the quality of care in clinical settings and the research on the cancer domain would be enhanced to a large extent.

Table 4: Mapping of each CRC core data element to FHIR resources

| Data element or subcategory	Description	Category	SCTID	Cardinality	FHIR resource (item)	Type
Name	Name of patient	Administrative	734841007	0.*	Patient (name)	Human name (text)
Gender	Male	Administrative	248153007	0.1	Patient (gender)	Code
Date of diagnosis	Date of first diagnosis in record	Diagnosis	-------------	0.1	Condition (recorded date)	date
Drug name	Capecitabine	Drug	386906001	1.1	Medication request (medication)	Codeable reference
Type of recurrence	Local	Follow-up	255470001	----	Schedule (Extension), also coded in diagnostic report	Extension (coded)
Date of imaging	Date of imaging performed	Diagnosis	410672004	0.1	Procedure (occurrence date time)	Date
Surgery	Left hemicolectomy	Diagnosis	82619000	0.1	Specimen (processing method)	Codable concept
Smoking	History of smoking of the patient	History of conditions	77176002	0.1	Condition (code)	Codable concept
Breast cancer	Family history of breast cancer of the patient	History of conditions	254837009	1.1	Family member history (code)	Codable concept
General symptom	Fatigue	Signs and symptoms	84229001	1.1	Observation (code)	Codable concept
Gastrointestinal symptom	Hematochezia	Signs and symptoms	405729008	1.1	Observation (code)	Codable concept
Gastrointestinal symptom	Rectal bleeding	Signs and symptoms	50960005	1.1	Observation (code)	Codable concept
Radiotherapy type	Interstitial brachytherapy	Radiotherapy	113120007	0.1	Procedure (code)	Codable concept
Surgery method	Open surgery	Surgery	129236007	0.*	Procedure (category)	Codable concept
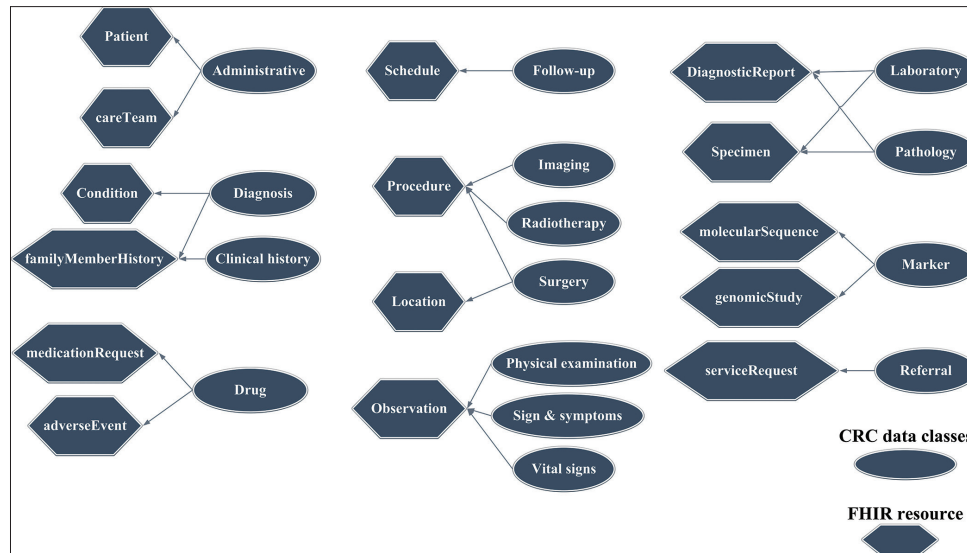
Figure 3: The general schema of the mapping of CRC data categories

Some points considered in this study included using the literature review and eliciting opinions of expert the panel of experts on CRC core data elements to obtain MDS for CRC. Developing the MDS before the mapping process at the national level is beneficial in standardizing core data elements. Collecting a minimum data set eliminates the need for various mapping scenarios using various forms, which is crucial in facilitating mapping process and achieving interoperability goals.

Using the SNOMED-CT before the mapping process has an advantageous impact on the perception of each core data element's terms and increases the accuracy of the mapping process to some extent. Standardizing the core data elements using the FHIR resources, considering the comprehensiveness of this standard, is beneficial in the cancer domain which was performed for the first time in this study in Iran. Due to using the codable concepts associated with most core data elements, the semantic interoperability of most CRC core data elements is achievable. The information modeling in this way has a potential for various advanced digital health tools, such as smartphones, tablets, and portable computers, to convey CRC data for efficient communication purposes. In Iran, the CDEs were considered a standard solution to data standardization. Indeed, it is not a correct option because the CDEs are not adapted to common terminologies without meta-data. So, misinterpretation and misperception of data would negatively impact on various ISs regarding the interoperability.

So far, few studies have been performed on standardizing the core data elements in the cancer domain using FHIR. Zong *et al*.[46] attempted to develop the standardized contents of case report forms in CRC clinical trial-based and EHR using the FHIR standard. They concluded that 1037 reports were successfully mapped to FHIR resources with an accuracy of 0.99 using the 200 randomly selected cases. Based on the results of the study, the FHIR was considered a suitable solution for the automatic population of case report forms. The CRC MDS mapped most of our core data elements to FHIR resources in the current research. The results of mapping the CRC core data elements indicated satisfactory resource adaptability. So, determining the MDS and then mapping performed in this study, or evaluating using different forms show the comprehensiveness of the FHIR resources. Also, in this study, using comprehensive terminologies such as SNOMED-CT significantly affected the apprehension of each CRC core data element concept in locating the relevant resource items. He-Xin *et al*.[47] introduced an integration solution to esophageal cancer screening cohorts by standardizing core data elements using FHIR and clinical data acquisition standards harmonization (CDASH) models. In their study, eight screening cohorts on esophageal cancer were utilized. The results of mapping in this model showed that this model can provide an effective and newly established data standard. This study deals with a high generalizability of the FHIR resources that we found in the current study.

Gonz'alez-Castro mapped the colon and breast cancer survivors' core data elements to FHIR resources as a valuable approach for data standardization. Two types of core data elements associated with these cancers were leveraged: clinical patient data, including general and medical history, condition pertained to cancer, tumor marker, comorbidities, treatment, clinical encounters, and patient-generated data such as patient-reported outcome measures and sensing data. Their study results showed that the FHIR-based cancer data model

could facilitate the prospective and retrospective data exchangeability and evince the secondary use of the real-world data. Also, they cited that common data elements (CDEs) are the constrained approach for data standardization due to incompatibility to uniform data standards and terminologies.[48] Developing an MDS is a fundamental step toward data harmonization, and we need more technical requirements for data integration in today's sophisticated healthcare and clinical trial settings.

Osterman *et al*.[49] developed an enhancing way of uniformity in oncology core data elements. The minimal CDEs were extracted by specialists' opinions in specific domains such as informatics, oncologists, and standardization and terminological fields. The standard cancer data classes comprised patient characteristics, laboratory and vital information, disease-specific, genomics profiles, and treatment and outcome information, covering 90 core data elements. The high conceptual level of the oncology core data elements by mapping process was performed in version 1 of the standardized data and introduced in the study. The FHIR-based cancer information model was agreed upon by 86.5% of the panel of experts. By consulting with experts, they tried to augment this agreement with more revisions in the next model version. Their study showed that the cancer information model significantly promotes the quality of care and research in the cancer domain by establishing a suitable infrastructure of data exchangeability.

More concisely, other studies utilized the FHIR modeling for data standardization approach in various cancer domains such as pan-oncology cohorts on genetic and lab tests,[50] precision medicine,[51] deep phenotype,[45] distributed cancer research,[52] or in the cancer-specific domain such as esophageal.[47] So far, no study has been conducted on using the FHIR resource to harmonize the Iran's cancer domain core data elements. The cancer data standardization in our country has been done at the level of developing CDE; for example, such study has been conducted on thyroid cancer,[53] breast cancer,[34] and information management of cancer[44] as a comprehensive approach for gathering, storing, retrieving, and sharing health data in clinical ISs. So, this is the first step forward in harmonizing the scattered core data elements in CRC by mapping to the FHIR resources in Iran.

While performing this study, we confronted some concerns that should be noted: We used 30 expert panel of experts to determine the most common CRC core data elements. So, this number of specialists may affect the Delphi technique results, generalizability, and final CRC core data elements. We suggest more experts for future studies in this respect. The mapping process is just performed from CRC core data elements to FHIR resources in this study. It is expected to consider mapping more cancer domains, such as breast and gastrointestinal organs, for subsequent research in this domain. Also, there are some limitations regarding the FHIR standard: for some core data elements, we could attribute them to multiple resources, even for elements that had precise definitions. For example, for a care team member, we could map it to practitioner and related person resources. Although consulting with experts and using various thesauri assisted us in better perception of terms, this reality always exists in FHIR standards that should be considered. Also, the semantic interoperability of the FHIR is exposed to challenges. In this respect, we could define several resources for similar meanings in FHIR resources. For example, suppose a person does not have a problem concerning drinking alcohol. In that case, we can code SNOMED-CT as a non-drinker for observation as a codable concept or attribute it to a problem drinker as a codable concept in a similar resource or add the value code of no as a modifier. Another limitation of this standard is that the details of the core data elements and all aspects of the terms meaning, such as the changeability, are not considered. In this study, we tried to use the extensions to embed more details of core data elements in the resource. Using an extension may affect the interoperability of the FHIR-based information models, and we suggest not using them as much as possible. However, some intrinsic constraints associated with FHIR standards, as mentioned in this study, exist, and future studies should consider these challenges when building interoperable content using the FHIR standards.

Conclusion

This study showed a new approach to standardizing the core data elements of CRC in Iran. Using the FHIR-based information model we built in this study assures the uniformness of CRC data and facilitates data sharing between various institutions. By using this standard to standardize the core data elements for CRC, the effectiveness and efficiency of healthcare data interoperability between different healthcare ISs, especially, the EHRs would be enhanced. It will lead to increased communication of various healthcare service providers such as physicians in different healthcare settings. By promoting data exchangeability between various healthcare providers and increasing the suitable use of healthcare data, the use of evidence-based approaches in clinical practices would be increased, leading to promoting patients' health outcomes. Increasing the use of healthcare data in this way also leads to better actions based on the evidence-based approach at the community level. For example, health policymakers and managers can leverage these data

to better allocate healthcare resources and funds to healthcare facilities more efficiently at the macro level. The healthcare managers can arrange the health resources and services based on the data that can be obtained by integrated HISs in various healthcare environments. The more widely integrated healthcare data and consequently increased secondary use of them also promote the research on the public health domain and have a crucial role in advancing the research practice on promoting the health status of people. Governments can improve the availability of high-quality healthcare by adjusting the politics in various aspects such as medical services, drugs, and medical insurance based on enhanced use of healthcare data and evidence-based decisions at the community level.

Ethics approval

Acknowledgments

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

References

1. Alzahrani SM, Al Doghaither HA, Al-Ghafari AB. General insight into cancer: An overview of colorectal cancer. Mol Clin Oncol 2021;15:1-8. doi: 10.3892/mco.2021.2433.

2. Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi REM, Corcione F. Worldwide burden of colorectal cancer: A review. Updates Surg 2016;68:7-11.

3. Brown KGM, Solomon MJ, Mahon K, O'Shannassy S. Management of colorectal cancer. BMJ 2019;366:l4561. doi: 10.1136/bmj.l4561.

4. Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: A review. JAMA 2021;325:669-85.

5. Vekic B, Dragojevic-Simic V, Jakovljevic M, Kalezic M, Zagorac Z, Dragovic S, *et al*. A correlation study of the colorectal cancer statistics and economic indicators in selected Balkan countries. Front Public Health 2020;8:29. doi: 10.3389/fpubh.2020.00029.

6. Al-Zalabani A. Preventability of colorectal cancer in Saudi Arabia: Fraction of cases attributable to modifiable risk factors in 2015–2040. Int J Environ Res Public Health 2020;17:320. doi: 10.3390/ijerph17010320.

7. Li N, Lu B, Luo C, Cai J, Lu M, Zhang Y, *et al*. Incidence, mortality, survival, risk factor and screening of colorectal cancer: A comparison among China, Europe, and northern America. Cancer Lett 2021;522:255-68.

8. Lewandowska A, Rudzki G, Lewandowski T, Stryjkowska-Góra A, Rudzki S. Risk factors for the diagnosis of colorectal cancer. Cancer Control 2022;29:10732748211056692. doi: 10.1177/10732748211056692.

9. Kolligs FT. Diagnostics and epidemiology of colorectal cancer. Viszeralmedizin 2016;32:158-64.

10. Sawicki T, Ruszkowska M, Danielewicz A, Niedźwiedzka E, Arłukowicz T, Przybyłowicz KE. A review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. Cancers 2021;13:2025. doi: 10.3390/cancers13092025.

11. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, *et al*. Colorectal cancer statistics, 2020. CA Cancer J Clin 2020;70:145-64.

12. Marley AR, Nan H. Epidemiology of colorectal cancer. Int J Mol Epidemiol Genet 2016;7:105-14.

13. Melina A, Mónica SS, Mathieu L, Isabelle S, Ahmedin J, Freddie B. Global patterns and trends in colorectal cancer incidence and mortality. Gut 2017;66:683-91.

14. Maheri M, Darabi F, Khalili S, Hesari M. The effect of educational intervention based on theory of planned behavior on the improvement of colorectal cancer screening intention among average-risk individuals referring to Asadabad city clinics. J Educ Health Promot. 2021;10:57.

15. Khazaei S, Salmani F, Moodi M. Evaluation of health belief model-based educational intervention on colorectal cancer screening behavior at South Khorasan, Iran. J Educ Health Promot. 2022;11:52.

16. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol 2021;14:101174. doi: 10.1016/j.tranon.2021.101174.

17. Keum N, Giovannucci E. Global burden of colorectal cancer: Emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol 2019;16:713-32.

18. Liu J, Li C, Xu J, Wu H. A patient-oriented clinical decision support system for CRC risk assessment and preventative care. BMC Med Inform Decis Mak 2018;18:118.

19. Condry MW, Quan XI. Digital health innovation, informatics opportunity, and challenges. IEEE Eng Manag Rev 2021;49:81-8.

20. Sodergren SC, Wheelwright SJ, Permyakova NV, Patel M, Calman L, Smith PWF, *et al*. Supportive care needs of patients following treatment for colorectal cancer: Risk factors for unmet needs and the association between unmet needs and health-related quality of life—results from the ColoREctal Wellbeing (CREW) study. J Cancer Survivorship 2019;13:899-909.

21. Sirintrapun SJ, Artz DR. Health information systems. Clin Lab Med 2016;36:133-52.

22. Sligo J, Gauld R, Roberts V, Villa L. A literature review for large-scale health information system project planning, implementation and evaluation. Int J Med Inform 2017;97:86-97.

23. Safdari R, Ghazi Saeedi MG, Masoumi-Asl H, Rezaei-Hachesu P, Mirnia K, Mohammadzadeh N, *et al*. National minimum data set for antimicrobial resistance management: Toward global surveillance system. Iranian journal of medical sciences. 2018;43(5):494-505.

24. Collen MF, Hammond WE. Development of medical information systems (MISs). The History of Medical Informatics in the United States. UK: Springer, London; 2015. p. 123-206.

25. Job Van B, Sarah L, Gary P, Katia V. Towards optimum reporting of pulmonary effectiveness databases and outcomes (TORPEDO): A modified Delphi exercise to identify a core dataset for asthma and COPD studies. Eur Respir J 2020;56(Suppl 64):2595. doi: 10.1183/13993003.congress-2020.2595.

26. Teixeira L, Cardoso I, Oliveira e Sá J, Madeira F. Are health information systems ready for the digital transformation in Portugal? Challenges and future perspectives. Healthcare 2023;11:712. doi: 10.3390/healthcare11050712.

27. Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The fast health interoperability resources (FHIR) standard: Systematic literature review of implementations, applications,

challenges and opportunities. JMIR Med Inform 2021;9:e21929. doi: 10.2196/21929.

28. Guo H, Scriney M, Liu K. An ostensive information architecture to enhance semantic interoperability for healthcare information systems. Inf Syst Front 2023. doi: 10.1007/s10796-023-10379-5.

29. Jahanbakhsh M, Sharifi M, Ayat M. The status of hospital information systems in Iranian hospitals. Acta Inform Med 2014;22:268.

30. Hong N, Wen A, Shen F, Sohn S, Wang C, Liu H, *et al*. Developing a scalable FHIR-based clinical data normalization pipeline for standardizing and integrating unstructured and structured electronic health record data. JAMIA Open 2019;2:570-9.

31. Montazeri M, Khajouei R, Mahdavi A, Ahmadian L. Developing a minimum data set for cardiovascular computerized physician order entry (CPOE) and mapping the data set to FHIR: A multi-method approach. J Med Syst 2023;47:47.

32. Mercorella M, Ciampi M, Esposito M, Esposito A, Pietro GD, editors. An Architectural Model for Extracting FHIR Resources from CDA Documents. 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Naples, Italy, 2016 28 Nov.-1 Dec. 2016.

33. Mirbagheri E, Ahmadi M, Salmanian S. Common data elements of breast cancer for research databases: A systematic review. J Family Med Prim Care 2020;9:1296-301.

34. Ghaneie M, Rezaie A, Ghorbani N, Heidari R, Arjomandi M, Zare M. Designing a minimum data set for breast cancer: A starting point for breast cancer registration in Iran. Iran J Public Health 2013;42(Suppl 1):66-73.

35. Bitarafan B, Jahanshahi A, Shahbazian H, Nooshabadi Z. Designing a minimum data set for thyroid cancer in Iran. Int J Cancer Manag 2020;13:e96997. doi: 10.5812/ijcm.96997.

36. Bernardi FA, Mello de Oliveira B, Bettiol Yamada D, Artifon M, Schmidt AM, Machado Scheibe V, Alves D, Felix TM. The minimum data set for rare diseases: Systematic review. J Med Internet Res. 2023;25:e44641.

37. Lee SH, Song JH, Kim IK, Kim JW. Clinical document architecture integration system to support patient referral and reply letters. Health Inform J. 2016;22:160-70.

38. Goenaga I, Lahuerta X, Atutxa A, Gojenola K. A section identification tool: Towards hl7 cda/ccr standardization in spanish discharge summaries. J Biomed Inform. 2021;121:103875.

39. Direito B, Santos A, Mouga S, Lima J, Brás P, Oliveira G, Castelo-Branco M. Design and Implementation of a Collaborative Clinical Practice and Research Documentation System Using SNOMED-CT and HL7-CDA in the Context of a Pediatric Neurodevelopmental Unit. Healthcare. 2023;11:973. doi: 10.3390/healthcare11070973.

40. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform 2019;94:103188. doi: 10.1016/j.jbi.2019.103188.

41. Savoska S, Ristevski B, Trajkovik V. Personal health record data-driven integration of heterogeneous data. Data-Driven Approach for Bio-medical and Healthcare. Singapore: Springer; 2022. p. 1-21.

42. Jaffe C, Nguyen V, Kubick WR, Cooper T, Leftwich RB, Hammond WE. Standards in biomedical informatics. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. Switzerland: Springer; 2021. p. 205-40.

43. Khan WA, Khattak AM, Hussain M, Amin MB, Afzal M, Nugent C, *et al*. An adaptive semantic based mediation system for data interoperability among health information systems. J Med Syst 2014;38:1-18. doi: 10.1007/s10916-014-0028-y.

44. Sadoughi F, Safdari R, Meraji M, Ramzan Ghorbani N, Ghazisaeedi M. Designing a minimum data set for national cancer registery in Iran. Nurs Midwifery J 2013;11:622-9.

45. Hochheiser H, Castine M, Harris D, Savova G, Jacobson RS. An information model for computable cancer phenotypes. BMC Med Inform Decis Mak 2016;16:121.

46. Zong N, Wen A, Stone DJ, Sharma DK, Wang C, Yu Y, *et al*. Developing an FHIR-based computational pipeline for automatic population of case report forms for colorectal cancer clinical trials using electronic health records. JCO Clin Cancer Inform 2020;4:201-9.

47. He-xin Y, Lu-ting G, Yong-le Z, Xuan L, Yi-fang Z, Feng B, *et al*. Establishment of technical specifications for data sharing in esophageal cancer screening cohort based on FHIR model and CDASH model. Chinese J Disease Control Prev 2021;25:1198.

48. González-Castro L, Cal-González VM, Del Fiol G, López-Nores M. CASIDE: A data model for interoperable cancer survivorship information based on FHIR. J Biomed Inform 2021;124:103953. doi: 10.1016/j.jbi.2021.103953.

49. Osterman TJ, Terry M, Miller RS. Improving cancer data interoperability: The promise of the minimal common oncology data elements (mCODE) initiative. JCO Clin Cancer Inform 2020;4:993-1001.

50. Zong N, Sharma DK, Yu Y, Egan JB, Davila JI, Wang C, *et al*. Developing a FHIR-based framework for phenome wide association studies: A case study with a pan-cancer cohort. AMIA Joint Summits on Translational Science proceedings AMIA Jt Summits Transl Sci 2020;2020:750-9.

51. Alterovitz G, Heale B, Jones J, Kreda D, Lin F, Liu L, *et al*. FHIR Genomics: Enabling standardization for precision medicine use cases. NPJ Genom Med 2020;5:13.

52. Lambarki M, Kern J, Croft D, Engels C, Deppenwiese N, Kerscher A, *et al*. Oncology on FHIR: A data model for distributed cancer research. Stud Health Technol Inform 2021;278:203-10.

53. Bitarafan B, Jahanshahi A, Shahbazian H, Nooshabadi Z. Designing a minimum data set for thyroid cancer in Iran. Int J Cancer Manag 2020;131-6.