Original Article

Development of Information Roadmap for Personalized Colorectal Cancer Screening in Iran

Elham Maserat[1*], Zeinab Mohammadzadeh[2], Mohammad Reza Zali[3]

[1]Department of Medical Informatics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran
[2]Health Information Technology Department, School of Management & Medical Informatics, Tabriz University of Medical Sciences, Tabriz, Iran
[3]Research Institute of Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Science, Tehran, Iran

ABSTRACT

Background:
A data-driven colorectal cancer screening strategy based on a personalized approach can improve health outcomes. This study aims to develop an information road map for personalized colorectal cancer screening in Iran.

Materials and Methods:
This study is a mix-method research (MMR) which consisted of three phases: phase I, the development of a checklist with 275 items for assessing required data elements of personalized colorectal cancer screening; phase II, situational analysis of colorectal cancer screening dataset according to the checklist; phase III, development of national information roadmap for personalized colorectal cancer screening with in-depth interview and focus groups.

Results:
Personalized datasets of colorectal cancer screening were defined in four dimensions, including a clinical dataset, a genetic dataset, a demographic dataset, and a social determinant dataset. In the next step data elements of colorectal cancer screening based on personalized datasets were analyzed. Of the 275 items, only 96 items are recorded. In the final step, a national information roadmap of personalized colorectal cancer screening with 6 levels was developed.

Conclusion:
Personalized screening based on integration dataset play a key role in the successful implementation of the screening program. Implementation of a national roadmap can assist to improve the quality of data in personalized screening.

Keywords: Colorectal cancer, Screening strategy, Information road map, personalized datasets

Please cite this paper as:
Maserat E, Mohammadzadeh Z, Zali MR. Development of information roadmap for personalized colorectal cancer screening in Iran. *Govaresh* 2022;27: 262-272.

Corresponding author:
Elham Maserat, PhD of Health Information Managment
Address: Tehran,Jalal AleAhmad, Nasr
P.O.Box: 14115-111
Tel: +98 9124866896
Telefax: +98 2884530
Email: e.maserat@modares.ac.ir

Received: 9 May 2022
Edited: 17 Dec. 2022
Accepted: 18 Dec. 2022

INTRODUCTION

Colorectal cancer (CRC) is the fourth leading cause of cancer mortality worldwide (1). The global burden of CRC is expected to increase by 60% to more than 2.2 million new CRC cases and 1.1 million deaths by 2030 (1). According to the GLOBOCAN, 2012 v1.0, the incidence of CRC will double by 2030 in Iran (2). The economic burden of CRC in Iran was estimated at US$298,148,718 in 2012. Considering the high economic burden and mortality rate of CRC on healthcare organizations, policies should be approved to reduce the disease and increase the prevention and early diagnosis of CRC (3). Studies show that CRC is one of the most preventable cancers if detected early (4). CRC incidence and mortality are reduced by regular screening (5-8). Also, screening can improve patient safety and ultimately reduce healthcare costs (9).

The decision-making approach for screening programs has become more complex in recent decades (10,11). CRC is heterogeneous cancer that is caused by multiple risk factors (12,13). Tumor heterogeneity as the greatest challenge must be considerable (14). Thus, customizing and personalizing decision-making for screening is recommended (15). Although there is an increasing number of screening strategies using molecular technologies, all of them do not have personalized criteria for screening approaches (16). Personalized medicine aims to offer and design appropriate diagnoses and treatments by individual patients' characteristics (17,18).

Personalized medicine focused on the integration of genomics and clinical dataset for supporting prevention strategies (19). Developing CRC personalized prevention could lead to the more effective usage of health resources (20,21). Also, personalized cancer care as with an individual clinical assessment approach can minimize cost and reduce efficacy (22). Personalized and patient-specific screening schedules facilitate the early stratification of at-risk individuals and detect significant biomarkers for predicting clinical status in individual patients (23). While individualized screening is an affordable strategy, there are challenges related to implementing personalized CRC screening (24). One of the big challenges facing personalized strategies is a lack of data (25).

The complement of data in risk assessment of precise cancer screening is an important factor (26). Also, the meaningful use of data for personalized protocols is essential. Integration and precise interpretation of the massive amount of data play a vital role in empowering personalized medicine (18). A data-driven cancer screening strategy is based on a personalized approach to improve health outcomes and manage healthcare costs (27). The CRC screening process involves collecting and analyzing a massive volume of clinical data for selecting appropriate evidence-based interventions (28). Therefore, the lack of clinical data as a barrier to implementing the personalized protocol is considerable (29). Panahiazar and coauthors explored some of the challenges in using data in personalized strategies, including a variation of the data, the quality of the data, and the volume and velocity of the data (18). We need vast amounts of data (clinical, environmental, and genetic datasets) for personalized programs. These huge datasets are extracted from different and heterogeneous sources. The integration of these data elements is a core barrier to personalized programs (30). One of the other challenges is the quality of data. Completeness and quality of data are very important factors in the decision-making process of cancer prevention (16). The genetic and clinical data reports of colorectal cancer include various and extensive components (31). In this paper, the national personalized datasets for CRC screening were developed and current datasets of CRC screening according to the personalized format were analyzed. Also, to address the above challenges, in our present study, we explored an information roadmap for personalized CRC screening. The information roadmap can describe the relationship between multiple sources and heterogeneous data components. Generally, the information roadmap optimizes integration of heterogeneous process (32). With regard to the importance of this subject, our study developed information roadmap for personalized CRC screening in Iran. In this project, we must fulfill the requirements set out by the availability of the screening documents in the research institute. We aimed to develop a personalized screening roadmap for integrating of the heterologous datasets.

MATERIALS AND METHODS

The methodology used in this paper is mixed method

research (MMR), which combines quantitative and qualitative methodology. This project was conducted at the Research Institute of Gastroenterology and Liver Disease (RIGLD), Shahid Beheshti University in Iran. The comprehensive plan for the screening of CRC has been designed by RIGLD for the last 18 years (33). This project was performed from 2016 to 2017.

In the first step, a 287-item Checklist 1: (yes, no) was developed for assessing required data elements for personalized screening. This checklist contained four dimensions sourced from the literature. In this step, all books, articles, research projects, thesis, manual and scientific reports were extracted from MEDLINE, The Institute of Electrical and Electronics Engineers (IEEE), Scholar, Web of Sciences, Scopus, ProQuest, and databases related to personalized CRC screening. We synthesized reliable evidence from multiple sources for determining personalized datasets of this checklist. The content validity of the developed checklist was assessed based on literature reviews and opinions of the experts related to the CRC screening program. The descriptive analyses were performed by SPSS software version 24. The status of colorectal cancer screening information system was investigated in the institute

The second step of this paper was a qualitative study. A National information roadmap of personalized CRC screening was developed in this step. A literature review was conducted for identifying components of the information roadmap. Then both in-depth interviews and focus group discussions were performed by clinicians and technical experts in CRC screening. In this study, 37 experts from 15 different fields related to CRC screening participated in the interview. The participants consisted of three epidemiologists, six genetic experts, one biochemist, one molecular biologist, two pathologists, one oncologist, seven gastroenterologists, five general medicine experts, one anesthesiologist, one surgeon, one psychometrist, two social medicine experts, three nutritionists, two internal medicine specialists, and one statistician. The researchers explained the study and obtained initial consent for further contact from participants. Also, the researcher asked for consent to audio-record the interviews. Semi-structured interview with six major themes (personalized data set, data architecture, data integration, data descriptor, monitoring program, and screening program) was used. Participants' experiences about components of the information roadmap were collected by interview. We used content analysis and frequency distributions for data analysis. For the final approval of the information roadmap, focus groups with semi-structured discussions were conducted. 18 experts with 8 different fields related to CRC screening & information technology participated in the focus group discussion. The participants consisted of two information technologists, three software engineers, three statisticians, six genetic experts, one oncologist, one gastroenterologist, and one pathologist. This team had executive experience in the screening program for at least 5 years. Two focus groups were conducted after an in-depth interview to approve the final roadmap.

RESULTS

This article presented significant dimensions in three parts. The first part, national personalized datasets for screening was described. In the second part, data components of present CRC screening datasets based on approved datasets of personalized screening were assessed. In the third part, the information road map of personalized CRC screening was developed.

National datasets of personalized CRC screening

Personalized datasets of CRC screening were defined in four different dimensions containing clinical dataset (5 sub-dimensions, 162 items), genetic dataset (2 sub-dimensions, 67 items), demographic dataset (1 sub-dimension, 6 items), and social determinant dataset (3 sub-dimensions, 40 items). The next step was developing national personalized datasets.

Analysis of present CRC screening datasets based on the developed national personalized dataset

In this step, the data elements of CRC screening based on personalized datasets were analyzed. The clinical dataset contained five items of pathology, surgery, clinical history, colonoscopy, and signs & symptoms. The Colonoscopy dataset as a sub-item of clinical datasets contained seven subgroups, including patients' history for colonoscopy readiness, anesthesia risk assessment, quality of bowel preparation, the aim of colonoscopy,

technical procedure, colonoscopy results, and follow-up. The highest data elements of clinical dimension were related to the pathological dataset (53.6%) in the RIGLD screening program (Table 1). The lowest data elements of the clinical dimension were related to the clinical history dataset (3.4%). 73% of the pedigree data element was entered. Table 2 was illustrated a situational analysis of demographic & social determinant datasets based on personalized CRC screening. Datasets related to the perspective of the participant were considered as part of

Table 1. Situational analysis of Clinical dataset based on personalized CRC screening (34,35,36)

Dimensions	Clinical Datasets		Present system.n (%)
Colonoscopy	**Items**		
	Patient history for colonoscopy readiness	Taking anticoagulant drugs, Need for antibiotic prophylaxis, Existence of the pacemaker, Existence of implantable cardioverter-defibrillator, Rectal bleeding	1/5 (20)
	Anesthesiology risk-assessment	Non-disorder, Mild disorder, Sever disorder	0/1 (0)
	Quality of bowel preparation	Excellent, good, adequate, fair, poor, and inadequate preparation	0/1 (0)
	Aim of colonoscopy	Screening, Evaluation of abnormal results & monitoring the history of neoplasia or previous polyps, Monitoring of Crohn's disease and ulcerative colitis, Biopsy	0/4 (0)
	Technical procedure	Date and time of procedure, CPT modifier codes, Type of anesthetic drug with a dosage, Expert type of anesthetic drug prescriber, Level of anesthesia, Type of device used, model and number of tools, Monitoring of equipment, Causes of colonoscopy cancellation, Colonoscopy complication	1/10 (10)
	Colonoscopy Results	Accuracy of the last colonoscopy, Date of last colonoscopy (year and month), Cause and suggested interval for the next colonoscopy, Topology /size/shape, and number of polyps, Topology/ size, and other tumor descriptors, Type and number of biopsies, Suspected diagnoses of mucosal disorders (ulcerative colitis, Crohn's disease, ischemia, infection), Topology and type of mucosal-vascular disorders, Other diagnoses (hemorrhoids, Diverticulosis)	3/9 (33.3)
	Follow up	Unplanned and emergency intervention after colonoscopy, Review of pathologic result	0/2(0)
Pathology	Sample dates, pathology result dates, Sample ID, Clinical diagnosis of sample, laboratory ID, Pathologist ID, Sampler ID, Tumor topology, Tumor morphology, Pathology laboratory city, Degree of tumor differentiation, Size of the largest tumor, Size of other tumors, Tumor behavior, Interval between margin and tumor, Closest margin to the tumor, Farthest margin, Distal margin removed, Proximal margin removed, TNM grading, Dukes grading, Metastasis diagnosis date, diagnosis method of metastasis, First metastasis site, Other metastasis sites, Sample number, Number of lymph nodes involved, Number of metastatic lymph nodes, Polyp number, Polyp shapes, Polyp size, Macroscopic description, Microscopic description, Preoperative diagnosis, Postoperative diagnosis, Date, and type of first tumor, Topology code of the final diagnosis, morphology code of the final diagnosis, Polyp histology type, Tumor site by grading, Pattern of developmental disorders		22/41 (53.6)
Surgery	First diagnosis date, Microscopic diagnosis method, Macroscopic diagnosis method, Operative date, Preoperative diagnosis, Postoperative diagnosis, Description of surgery, Surgery techniques, Surgery equipment, Cancellation cause of surgery, Procedure aim, Surgery type, Sample ID, number/ type/ size and site of the removed sample, anesthesia type, Anesthesiologist ID, Surgical technologist ID, Nursing Assistant ID, Surgeon ID, Cause of re-surgery, Colonoscopy result after surgery, Bilateral surgical procedures, Emergency surgery, Surgery time, Vital status after surgery, Death date, Cause of death, Type and number of palliative care before surgery, Type and number of palliative care after surgery, Duration of hospitalization due to surgical procedure, Cause of recurrent cancer, First date of recurrent cancer, Site of recurrent cancer, Recurrence number, Complication of surgery		12/35 (34.2)

Table 1. Continued.

Dimensions	Clinical Datasets	Present system.n (%)
Sign & Symptom	Changes in bowel movements (number of times and quality), Painful bowel movements, Blood (red blood or dark blood) in the stool, Rectum bleeding, Vomiting and nausea, Weight loss without cause, Diarrhea, Fatigue and weakness, Constipation, bowels are not completely emptied, General abdominal discomfort, Painful abdominal contractions, Fever with unknown cause, Anemia, Rectal pain, Indigestion, Itching around the anal, Depression, Digestive disorders, Incontinence in bowel function, Painful bowel movements, Signs of mass in the abdomen, Obstructive bowel symptoms, Signs of perforation of the bowel, chief complaint, and duration of its complaint	9/25 (36)
Clinical History	Inflammatory bowel disease, Diabetes, Abnormal blood pressure, Abnormal blood sugar, Abnormal cholesterol, Abnormal triglyceride, Abnormal hemoglobin, Cardiovascular disease, Surgery (especially appendix, stomach, gallbladder), Use of alternative hormonal treatments, Polyp, Cancer, History of cancer, Cancer type, Cancer site, Age of cancer diagnosis, History of clinical and screening procedure such as colonoscopy/ sigmoidoscopy/ barium enema, result of the clinical and screening procedure, Genetic and immunohistochemistry tests, History of first treatment such as surgery/ radiotherapy/ chemotherapy/ biopsy, Taking non-steroidal anti-inflammatory drugs, Taking aspirin, Respiratory disorders, Asthma, Musculoskeletal disorders, Mental and behavioral disorders, History of coronary artery disease, History of metastases (Type of cancer, age of diagnosis and metastasis site)	1/29(3.4)

Table 2. Situational analysis of demographic & social determinant dataset based on personalized CRC screening (37,38)

Target Population **Patient, Relatives with colorectal cancer, Non- colorectal cancer relatives, Non-patient relatives and Non-patient volunteers**		
Dimensions	**Demographic & Social determinant Datasets**	**Present system n (%)**
Demographic Datasets		
Basic Dataset	Patient name, National code, Patient ID, Blood type, Birth date, Sex	6/6 (100)
Social determinant Datasets		
Environmental Dataset	Ethnicity, Place of birth, Insurance status, Insurance number, Marital status, Education level, Major, Job status, Religion, Residence address, Postal code, Residence duration, Patient phone number, Relative phone number, Sleep status, Access to medical & prevention services, Income rate	13/17 (34.3)
Lifestyle Dataset	BMI, Waist circumference, Number and time of consuming vegetables and fruits during the week, Exposure to chemicals and other types of pollutants, Number of time smoking per day, Number of times smoking hookahs and tobacco, Number of times and rate of injecting drug use and other tobacco, Palliative drug use, Number of times and rate of fried food consumed during the week, Number of times and rate of consuming red meat during the week, Number of times and rate of physical activity during the week, Number of times and rate of consuming fiber during the week, Number of times and rate of consuming alcohol, pattern of consumption of alternative hormonal treatments, Safety lifestyle pattern, Menopause statues, Menstrual status	2/17 (11.7)
The perspective of Participant Dataset	Perceived susceptibility to the screening program, Perceived severity of the screening program, Perceived benefits of screening programs, Perceived barriers to screening, Perceived threat of screening programs, Perceived execution process of screening programs	0/6 (0)

the social determinants dataset. However, there were no data elements of the perspective of participants in the CRC screening program (Table 3).

Information roadmap for personalized colorectal cancer screening

We developed a national information roadmap for

Table 3. Situational analysis of genetic dataset based on personalized CRC screening (39,40)

Dimensions	Genetic Datasets	Present system n,(%)
Genome Dataset	Number and gender of suspected cases to HNPCC, Number and gender of suspected cases to FAP, Number and gender of suspected cases to AFAP, Number and gender of suspected cases to MAP, Number and gender of cases to HNPCC, Number and gender cases to FAP, Number and gender of cases to AFAP, Number and gender of cases to MAP, Number of pathologic blocks for genetic tests, Blood sample ID, IHC statues, MSI statues, Genetic laboratory ID, Genes expression status (MLH1, MSH2, MSH6, PMS2), Mutation status(MLH1, MSH2, MSH6, PMS2), Colorectal cancer genetic markers, Pattern of methylation of genes in colorectal cancer, Selective inhibitors of enzymes, Type of arrangement and penetration of immune cells in colorectal cancer, Mutations of pathogenesis of hereditary syndromes, Non-heritable pathogen mutations, Polymorphisms and mutations in expression of the oncogenic genes, LOH in colorectal cancer, Methylation of genes in the pathology block, Epigenetic data in colorectal cancer, LPS microbes in genes methylation, Expression of Carcinogenic Gene in Hyper Plastic Polyps, HLA-neoclassic expression to estimate metastasis, Data on changes in telomerase size to estimate survival, Stem cell markers in colorectal cancer, Expression of APC and B7 genes in colorectal polyps, Gene expression profiles related to Angiogenesis for estimating colorectal metastasis, Sequence of genes in the cancer pathway, Genetic mapping in cancer pathway, Gene therapy data on cancer control, Pattern of changes in the expression of genes in cancer progression, Pattern of changes in expression of inhibitor genes, nanoparticles data in treatment, Tumor suppressor proteins, Genetic pattern recurrence, Identifier/ Ontology/ Cell Components/ Biological Processes of Inhibiting and Suppressing Genes	7/41 (17)
Pedigree Dataset	Number of main patient households, Number of first degree relatives, Number of second degree relatives, First degree relatives' gender, Second degree relatives' gender, First degree relatives' age, Second degree relatives' age, Special relative status (twins, multiple paired), Disease type in first-degree relatives (except for cancer), Disease type in second-degree relatives (except for cancer), Number of first-degree relatives with colorectal cancer, Number of second-degree relatives with colorectal cancer, Type of other cancers in first-degree relatives, Type of other cancers in second-degree relatives, Age of Cancer diagnosis in first second degree relatives, Age of Cancer diagnosis in second degree relatives, Primary site of cancer in first-degree relatives, Primary site of cancer in second-degree relatives, Metastatic site of cancer in first degree relatives, Metastatic site of cancer in second degree relatives, Age of death in first degree relatives, Age of death in second degree relatives, Cause of death in first degree relatives, Cause of death in second degree relatives, Polyp history in first degree relatives, Polyp history in second degree relatives	19/26(73)

personalized CRC screening through in-depth interviews and focus group discussions (Table 4). This roadmap was approved by CRC experts. Information leadership is at the top of this roadmap (Figure 1). This roadmap has six layers, including information leadership, personalized dataset, data integration, data architecture, data descriptor, and screening program layers. Information leadership describes a process that leads to data resource management and information infrastructure organization (34). The second layer is a personalized dataset. The next layer was presented that personalized screening requires the integration of clinical dataset, genetic dataset, and demographic and social determinants dataset. Data modeling and designed architecture as one of the layers of the roadmap for interoperability of screening dataset are essential factors. Modifiers and descriptors of the dataset were defined in the fifth layer. Developing of vision, mission, and goals of personalized screening as the infrastructure of the roadmap is dependent on the integrated dataset. Information governance and integrated dataset can support continuous editing vision, mission, and goals of the screening plan by the experiences of experts. Also, program stewardship and information resources management are integral parts of the information roadmap of personalized CRC screening.

DISCUSSIONS

Personalized datasets of CRC screening were defined in four dimensions with 275 items. In the next step data elements of CRC screening based on personalized datasets

Table 4. In-depth interview and focus group discussion: Selected comments about the national information roadmap of personalized CRC screening

Methods	Comments	
In-depth interview & focus group discussion	Themes	Selected Comments
	Personalized Data Set	"In my opinion, personalized datasets should include clinical, genetic and lifestyle information." [Statistician] "HNPCC and FAP have a key role in genetic datasets." [genetic expert]
	Data Architecture	"Based on my five years of experience, I think it's important to consider genetic metadata for an information road map of personalized CRC screening." [Information technologist]
	Data Integration	"The use of standards such as Hl7, SNOMED, LOINC, DICOM, and UMLS seem essential for integration." [Information software engineer]
	Data modifiers and descriptors	"All data quality components must be considered in the roadmap." [Information software engineer]
	Monitoring Program	"All components of the Stewardship plan must be considered" [Executive member/gastroenterologist]
	Screening Program	"In my opinion, the role of information governance in formulating goals, vision, and missions of the personalized screening program is significant." [Health information manager]

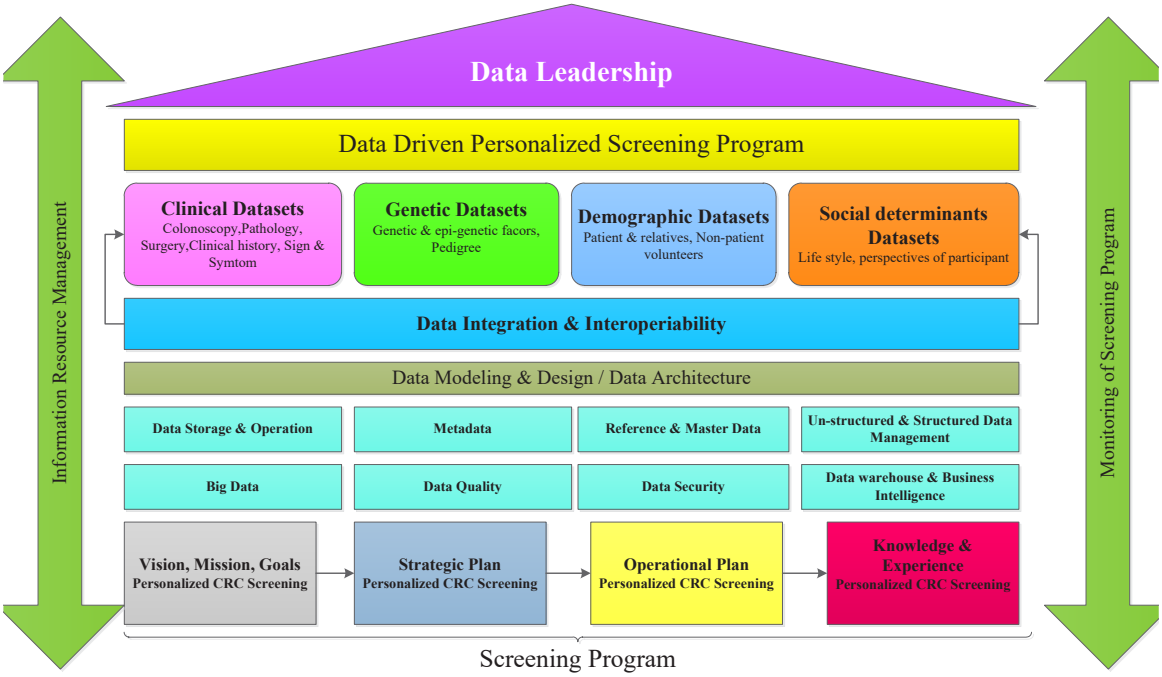

Figure 1. National information roadmap of personalized CRC screening

were analyzed. In the final step, the national information roadmap of personalized CRC screening with six layers was developed.

Personalized screening approaches can optimize the efficiency, equity, and safety of cancer screening, but will require precise and comprehensive patient information (41). Recent studies show that accurate and detailed information can support personalized prevention strategies (24,42,43). It has been emphasized that screening recommendations require comprehensive

patient information (41). Considering the vital role of patient information in effective and affordable screening, we developed national personalized datasets with 275 items. This personalized dataset has been approved by experts in 15 different fields related to CRC screening. Despite the importance of accurate and complete datasets in the more effective management of the personalized screening program, the existing datasets have limited and inadequate elements in our study. Of the 275 items, only 96 items are recorded. Only 17.8% of the clinical dataset of the screening program was entered, while clinical information plays a key role in risk assessment (44). Determining individual risk factors as significant elements of the successful implementation of personalizing CRC screening is considerable. Sufficient and up-to-date risk factors information is an integral part of more efficient individual screening (22). It is necessary to have a standard documentation tool for the evaluation of clinical datasets (45). We developed a standard checklist of the personalized dataset for the evaluation of the clinical dataset in this study. Detailed information is needed in the clinical part of the personalized dataset.

In this present study, only 9% of colonoscopy datasets were entered. Patient-centered colonoscopy has become a critical issue in CRC screening (46). The safety and effectiveness of the cancer screening program depend on the quality of the colonoscopy. For the high quality of this procedure, documentation of detailed and complete data is the significant parameter (47). In this research, the colonoscopy data was collected inadequately; more information needs to be added to the report.

Colonoscopy plays a key role in the screening process. The success of the screening program depends on the quality of the colonoscopy. Colonoscopy reports reflect the quality of the colonoscopy procedure (48). Colonoscopy reporting was poor in this study and the data collection process should be revised.

The highest data entry of clinical dimension was related to the pathological dataset in the screening program of RIGLD. According to studies, pathology sample is obtained in 30% to 50% of colonoscopy interventions (49). Accurate histopathological data is a requirement for providing high-quality care services to patients with CRC (49). Precise pathology reports can enhance screening recommendations for follow-up. Quality Assurance Task Group presents key data indicators for pathology documentation to achieve continuous quality improvement (CQI) (50).

The study proved that the highest data entry of genetic dimension was related to the pedigree dataset. Pedigree is useful for tracking and presenting detailed family history data (51-53). The Pedigree data element is a powerful dataset that can support genetic dimension in screening approaches (53).

The lifestyle factor is an important predictor for screening participants (54-58), while only 11.7% of lifestyle datasets are entered into the present system. Also, participants' perspective datasets, as a part of the social determinant dimension are not documented. In general, 62.5% of social determinant datasets were entered.

We need a standard dataset for comprehensive data documentation (59). Therefore, standard datasets for complete and accurate data gathering are one of the requirements of the personalized screening system. In this paper, the standard personalized dataset was developed and in the next step present datasets were assessed by this standard dataset and incomplete data elements were identified. Studies show that information gaps and deficiencies can affect the decision-making of clinicians (31,60,61). Thus, addressing the present dataset's deficiencies is necessary for better screening decision-making.

In addition, the completeness and accuracy of data are essential factors for increasing the effectiveness of the screening program. Clinical and genetic information should integrate for individual risk stratification in personalized prevention (62,63). To meet this need, we developed a national roadmap for personalized CRC screening. Data leadership is at the top of this roadmap. Information leadership as a layer of roadmap facilitates effective decision-making (34). Successful CRC screening depends on a precise and data-driven plan (64). In our information roadmap, there is continued interaction between personalized datasets and components of the screening plan such as the vision, missions, and goals of the program in the development roadmap. One of the caveats of this study is the fact that the data sample is related to one of the screening centers.

Personalized colorectal cancer screening is a comprehensive approach to prevention based on each person's unique datasets. One of the big challenges in this approach is the provision of complete and precise data. In this study, we developed a standard tool for information gap analysis. The results of this analysis can be considered in the screening program planning and quality improvement of documentation. The integrity of the high volume of information in a precise screening program is another problem. The developed roadmap identifies various data components of the program and integrates all of the information segments. It can be used as a tool for data process reengineering.

CONCLUSION

Due to the importance of integrated data in the personalized screening approach and the lack of such data sets, researchers conducted this project. Personalized prevention based on the integration dataset plays a key role in the more efficient implementation of the screening program. The developed roadmap can be used for the integration and interoperability of screening datasets. The efficiency of the decision-making process in the screening plan can be reduced by data deficiencies. Eliminating data deficiencies can improve the quality of documentation and may lead to improved screening performance. Therefore, the reason for data deficiencies and missing values should be identified and eliminated. In this study, information deficiencies were identified by a standard instrument. Entering data was inadequate and poor in the screening program. Implementation of a national roadmap can assist to improve the quality of data in personalized screening. According to recent studies, the use of standard datasets and indicators can help to identify information gaps and facilitate evidence base decision-making. Establishing an individual screening program requires a comprehensive and accurate database. In our study, data gaps were analyzed by a national checklist and a roadmap was developed for the interaction and integration of heterogeneous data.

In the future study, the present map will be implemented and the results will be reported. Continuous monitoring of the data process via this roadmap can facilitate quality improvement of personalized screening. Implementation of this map is expected to lead to comprehensive data for decision-making.

ACKNOWLEDGMENTS

The authors would like to thank the unsparing helps of RIGLD staff to collect useful data.

AUTHORS' CONTRIBUTION

EM: Design research, gathering data, analysis, and Interpretation of data
ZM: Design research, Analysis, and interpretation of data
MR Z: Design research, Analysis, and interpretation of data

AVAILABILITY OF DATA AND MATERIAL

The data that supported the findings of this study are available from the corresponding author on request

CONFLICT OF INTEREST

No conflict of interest was declared.

CONSENT FOR PUBLICATION

Not applicable

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The ethical approval was accepted by RIGLD ethical committee.

FUNDING

This study was funded with the authors' own contributions.

REFERENCES

1. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019;51(1):76-87.
2. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015 Mar 1;136(5):E359-86
3. Vahdatimanesh Z, Zendehdel K, Kbari Sari AA, Farhan F, Nahvijou A, Delavari A, et al. Economic burden of colorectal cancer in Iran in 2012. Med J Islam Repub Iran. 2017;31:115.
4. Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW, García FA, et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. JAMA. 2016;315(23):2564-2575. 2016;315(23):2564-75.
5. Berland LL, Monticciolo DL, Flores EJ, Malak SF, Yee J, Dyer DS. Relationships Between Health Care Disparities and Coverage Policies for Breast, Colon, and Lung Cancer Screening. J Am Coll Radiol. 2019;16(4 Pt B):580-585.
6. Pourhoseingholi MA, Vahedi M, Moghimi-Dehkordi B, Pourhoseingholi A, Ghafarnejad F, Maserat E, et al. Burden of hospitalization for gastrointestinal tract cancer patients-Results from a cross-sectional study in Tehran. Asian Pac J

Cancer Prev. 2009;10(1):107-10.
7. Rogers CR, Matthews P, Xu L, Boucher K, Riley C, Huntington M, et al. Interventions for increasing colorectal cancer screening uptake among African-American men: A systematic review and meta-analysis. PloS One. 2020;15(9):e0238354.
8. Kahi CJ, Anderson JC. Colorectal Cancer Screening and Surveillance. Clinical Gastrointestinal Endoscopy: Elsevier; 2019. p. 392-401. e3.
9. Baumel M, White EC. Colo-rectal cancer screening system and method. Google Patents; 2008.
10. Keating NL, Pace LE. Breast cancer screening in 2018: time for shared decision making. JAMA. 2018;319(17):1814-5.
11. Adunlin G, Diaby V, Montero AJ, Xiao H. Multicriteria decision analysis in oncology. Health Expect. 2015;18(6):1812-26.
12. Murphy N, Ward HA, Jenab M, Rothwell JA, Boutron-Ruault M-C, Carbonnel F, et al. Heterogeneity of colorectal cancer risk factors by anatomical subsite in 10 European countries: a multinational cohort study. Clin Gastroenterol Hepatol. 2019;17(7):1323-31. e6.
13. Myte R, Gylling B, Häggström J, Häggström C, Zingmark C, Löfgren Burström A, et al. Metabolic factors and the risk of colorectal cancer by KRAS and BRAF mutation status. Int J Cancer. 2019;145(2):327-337.
14. Corey L, Valente A, Wade K. Personalized medicine in gynecologic cancer: fact or fiction?. Surg Oncol Clin N Am. 2020;29(1):105-113.
15. Magrath M, Yang E, Singal AG. Personalizing Colon Cancer Screening: Role of Age and Comorbid Conditions. Current Colorectal Cancer Reports. 2017;13(3):220-6.
16. Binefa G, Rodríguez-Moranta F, Teule À, Medina-Hayas M. Colorectal cancer: from prevention to personalized medicine. World J Gastroenterol. 2014;20(22):6786-808.
17. Jain KK, Jain KK. Textbook of personalized medicine. New York: Springer; 2009 Aug 29.
18. Panahiazar M, Taslimitehrani V, Jadhav A, Pathak J, editors. Empowering personalized medicine with big data and semantic web technology: promises, challenges, and use cases. Proc IEEE Int Conf Big Data. 2014 Oct;2014:790-795.
19. Abul-Husn NS, Kenny EE. Personalized Medicine and the Power of Electronic Health Records. Cell. 2019;177(1):58-69.
20. Fletcher RH. Personalized screening for colorectal cancer. Med Care. 2008;46(9):S5-S9.
21. Kuipers EJ, Spaander MC. Personalized screening for colorectal cancer. Nat Rev Gastroenterol Hepatol. 2018;15(7):391-392.
22. Schilsky RL. Personalized medicine in oncology: the future is now. Nat Rev Drug Discov. 2010;9(5):363-6.
23. Walker R, Mejia J, Lee JK, Pimiento JM, Malafa M, Giuliano AR, et al. Personalizing gastric cancer screening with predictive modeling of disease progression biomarkers. Appl Immunohistochem Mol Morphol. 2019;27(4):270-277.
24. Robertson DJ, Ladabaum U. Opportunities and Challenges in Moving From Current Guidelines to Personalized Colorectal Cancer Screening. Gastroenterology. 2019;156(4):904-17.
25. Horgan D, Jansen M, Leyens L, Lal JA, Sudbrak R, Hackenitz E, et al. An index of barriers for the implementation of personalised medicine and pharmacogenomics in Europe. Public Health Genomics. 2014;17(5-6):287-98.
26. Arrospide A, Forné C, Rué M, Torà N, Mar J, Baré M. An assessment of existing models for individualized breast cancer risk estimation in a screening program in Spain. BMC Cancer. 2013;13(1):587.
27. Sharma A, Hostetter J, Morrison J, Wang K, Siegel E. Focused Decision Support: a Data Mining Tool to Query the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial Dataset and Guide Screening Management for the Individual Patient. J Digit Imaging . 2016;29(2):160-4.
28. Tan SC. Low penetrance genetic polymorphisms as potential biomarkers for colorectal cancer predisposition. J Gene Med. 2018;20(4):e3010.
29. Schaffhausen J. What Precisely Is Precision Medicine? Trends Pharmacol Sci. 2017;38(1):1-2.
30. Vidal ME, Endris KM, Jozashoori S, Karim F, Palma G. Semantic Data Integration of Big Biomedical Data for Supporting Personalised Medicine. Current Trends in Semantic Web Technologies: Theory and Practice: Springer; 2019. p. 25-56.
31. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144-51.
32. Chan JO. Enterprise information systems strategy and planning. J Am Academy Business. 2005;6(2):148-53.
33. Moghimi-Dehkordi B, Maserat E, Zali MR. Prevention strategy of colon cancer at risk population in Iran. J Med Council Iran. 2010;28(4):444-9.
34. Huvila I. Towards information leadership. Aslib Journal of Information Management. 2014;66(6):663-77.
35. Casati B, Bjugn R. Structured electronic template for histopathology reporting on colorectal carcinoma resections: five-year follow-up shows sustainable long-term quality improvement. Arch Pathol Lab Med. 2012;136(6):652-6.
36. Ihnát P, Delongová P, Horáček J, Rudinská LI, Vávra P, Zonča P. The impact of standard protocol implementation on the quality of colorectal cancer pathology reporting. World J Surg. 2015;39(1):259-65.
37. Chen J, He H, Jin H, McAullay D, Williams G, Kelman C, editors. Identifying risk groups associated with colorectal cancer. Data Mining; 2006: Springer.
38. Mitchell R, Brewster D, Campbell H, Porteous M, Wyllie A, Bird C, et al. Accuracy of reporting of family history of colorectal cancer. Gut. 2004;53(2):291-5.
39. Zhou Y, Boardman L, Miller R. Genetic testing for young-

onset colorectal cancer: case report and evidence-based clinical guidelines. Radiol Oncol. 2010;44(1):57-61.
40. Yuan Y, Zhu L, Xu D, Haixing J, Sun Y, Ding PR, et al. The prevalence of germline mutations in Chinese colorectal cancer patients with mismatch repair deficiency. Am Soci Clin Oncol. 2018.
41. Selby K, Bartlett-Esquilant G, Cornuz J. Personalized cancer screening: helping primary care rise to the challenge. Public Health Rev.2018;39:4.
42. Portnoi T, Yala A, Schuster T, Barzilay R, Dontchos B, Lamb L, et al. Deep Learning Model to Assess Cancer Risk on the Basis of a Breast MR Image Alone. Am J Roentgenol.2019:1-7.
43. Han PK, Duarte CW, Daggett S, Siewers A, Killam B, Smith KA, et al. Effects of personalized colorectal cancer risk information on laypersons' interest in colorectal cancer screening: The importance of individual differences. Patient Educ Couns. 2015;98(10):1280-6.
44. Wang Y, Zhu G, Engel B. Health risk assessment of trihalomethanes in water treatment plants in Jiangsu Province, China. Ecotoxicol Environ Saf. 2019;170:346-354.
45. Hoseinpourfard M, Abbasi Dezfouli S, Ayoubian A, Izadi M, Mahjob MP. Hospital compliance with clinical documentation standards: A descriptive study in two Iranian teaching hospitals. Int J Hospit Res. 2012;1(2):121-5.
46. Sewitch MJ, Azalgara VM, Sing M. Screening indication associated with lower likelihood of minor adverse events in patients undergoing outpatient colonoscopy. Gastroenterol Nurs. 2018;41(2):159-64.
47. Rembacken B, Hassan C, Riemann J, Chilton A, Rutter M, Dumonceau J-M, et al. Quality in screening colonoscopy: position statement of the European Society of Gastrointestinal Endoscopy (ESGE). Endoscopy. 2012;44(10):957-68.
48. Lee TJ, Rutter MD, Blanks RG, Moss SM, Goddard AF, Chilton A, et al. Colonoscopy quality measures: experience from the NHS Bowel Cancer Screening Programme. Gut. 2012;61(7):1050-7.
49. Haugland HK, Casati B, Dørum LM, Bjugn R. Template reporting matters—a nationwide study on histopathology reporting on colorectal carcinoma resections. Human Pathol. 2011;42(1):36-40.
50. Lieberman D, Nadel M, Smith RA, Atkin W, Duggirala SB, Fletcher R, et al. Standardized colonoscopy reporting and data system: report of the Quality Assurance Task Group of the National Colorectal Cancer Roundtable. Gastrointestinal Endosc. 2007;65(6):757-66.
51. Rieger AK, Mansmann UR. A Bayesian scoring rule on clustered event data for familial risk assessment–An example from colorectal cancer screening. Biom J. 2018;60(1):115-27.
52. Li W, Zhao LZ, Ma DW, Wang DZ, Shi L, Wang HL, et al. Predicting the risk for colorectal cancer with personal characteristics and fecal immunochemical test. Medicine. 2018; 97(18):e0529.
53. Chiu SY-H, Chen L-S, Yen AM-F, Chen H-H. Population-based proband-oriented pedigree information system: application to hypertension with population-based screening data (KCIS No. 25). J Am Med Inform Assoc. 2012;19(1):102-10.
54. Larsen IK, Grotmol T, Almendingen K, Hoff G. Lifestyle characteristics among participants in a Norwegian colorectal cancer screening trial. Eur J Cancer Prev. 2006;15(1):10-9.
55. Sutton S, Wardle J, Taylor T, McCaffery K, Williamson S, Edwards R, et al. Predictors of attendance in the United Kingdom flexible sigmoidoscopy screening trial. J Med Scr. 2000;7(2):99-104.
56. Weissfeld JL, Ling BS, Schoen RE, Bresalier RS, Riley T, Prorok PC, et al. Adherence to repeat screening flexible sigmoidoscopy in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Cancer. 2002;94(10):2569-76.
57. Ioannou GN, Chapko MK, Dominitz JA. Predictors of colorectal cancer screening participation in the United States. Am J Gastroenterol. 2003;98(9):2082-91.
58. Slattery M, Kinney A, Levin T. Factors associated with colorectal cancer screening in a population-based study: the impact of gender, health care source, and time. Prev Med. 2004;38(3):276-83.
59. Meldolesi E, van Soest J, Dinapoli N, Dekker A, Damiani A, Gambacorta MA, et al. An umbrella protocol for standardized data collection (SDC) in rectal cancer: a prospective uniform naming and procedure convention to support personalized medicine. Radiother Oncol. 2014;112(1):59-62.
60. Wagholikar K, Sohn S, Wu S, Kaggal V, Buehler S, Greenes RA, et al. Workflow-based data reconciliation for clinical decision support: case of colorectal cancer screening and surveillance. AMIA Jt Summits Transl Sci Proc. 2013;2013:269-73.
61. Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing electronic note quality using the physician documentation quality instrument (PDQI-9). Appl Clin Inform. 2012;3(2):164-174.
62. Läll K, Lepamets M, Palover M, Esko T, Metspalu A, Tõnisson N, et al. Polygenic prediction of breast cancer: comparison of genetic predictors and implications for risk stratification. BMC Cancer. 2019;19(1):557.
63. Auer R, Selby K, Bulliard J-L, Nichita C, Dorta G, Ducros C, et al. Shared decision making in the colorectal cancer screening program in the canton of Vaud.Rev Med Suisse. 2015;11(496):2209-15.
64. Levin B, Smith RA, Feldman GE, Colditz GA, Fletcher RH, Nadel M, et al. Promoting early detection tests for colorectal carcinoma and adenomatous polyps: a framework for action: the strategic plan of the National Colorectal Cancer Roundtable. Cancer. 2002;95(8):1618-28.