

# Development of Information Road map for Personalized Colorectal Cancer Screening in Iran

elham maserat (✉ [elhammaserat@gmail.com](mailto:elhammaserat@gmail.com))

Tarbiat Modares University Faculty of Medical Sciences <https://orcid.org/0000-0003-1359-8024>

**Zeinab Mohammadzadeh**

Tabriz University of Medical Sciences

**Mohammad Reza Zali**

Shaheed Beheshti University of Medical Sciences Research Institute for Gastroenterology and Liver Diseases

---

## Research article

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

**Posted Date:** March 13th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-17085/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** A data-driven colorectal cancer screening strategy based on personalized approach can improve health outcomes, facilitate early stratification of at-risk patients and reduce health care costs. This study aims to develop an information road map for personalized colorectal cancer screening in Iran.

**Methods** This study is a Mix-Method Research (MMR) which consisted of three phases: phase I, 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 road map 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 clinical dataset (5 sub-dimensions, 162 items), genetic dataset (2 sub-dimensions, 67 items), demographic dataset (1 sub-dimension, 6 items) and a social determinant dataset (3 sub-dimensions, 40 items). The next step data elements of colorectal cancer screening based on personalized datasets were analyzed. Of the 275-items, only 96 items are recorded. Only 17.8% of clinical dataset of screening program were entered. The highest data elements of clinical dimension were related to pathological datasets (53.6%) in the present screening program. The lowest data elements of the clinical dimensions were related to the clinical history dataset (3.4%). 73% of pedigree data elements and 15.33% of social determinant datasets were entered. In the final step, a national information road map of personalized CRC screening with 6 layers (information leadership, personalized datasets, data integration, data architecture, data descriptor, and screening program layers) was developed.

**Conclusion** Personalized screening based on integration dataset play a key role for the successful implementation of the screening program. Eliminating data deficiencies can improve the quality of documentation and may lead to improved screening performance. Therefore, standard datasets and indicators can help to identify information gaps and facilitate precise decision-making. Entering data was inadequate and poor in this study. Implementation of national road map can assist to improve the quality of data in personalized screening.

## Background

Colorectal cancer (CRC) is the fourth leading cause of cancer mortality worldwide (1). The global burden of colorectal cancer (CRC) is expected to increase by 60% to more than 2.2 million new colorectal cancer 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. Concerning the high economic burden and mortality rate of colorectal cancer on the health care 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).

Colorectal cancer incidence and mortality are reduced by regular screening (5–8). Also, screening can improve patient safety and ultimately reduce health care costs (9).

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 challenges 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 the personalized criteria for screening approaches (16). Personalized medicine aims are to offer and design appropriate diagnosis and treatment by individual patient's 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 individual clinical assessment approach can minimize cost and reduce efficacy (22). Personalized and patient-specific screening schedule to facilitate 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 massive amount of data play a vital role in empowering personalized medicine (18). A data driven cancer screening strategy based on 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 of appropriate evidence-based interventions (28). Therefore, the lack of clinical data as barriers to implementing of 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, 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 extract from different and heterogeneous sources. The integration of these data elements is a core barriers of personalized programs (30). One of the other challenges is the quality of data. Completeness and quality of data is a very important factor in the decision-making process of cancer prevention (16). In this paper, the national personalized datasets for colorectal cancer 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, we explored an information road map for personalized colorectal cancer screening. Information road map can describe the relationship between multiple sources and heterogeneous data components. Generally, the information road map optimizes integration of heterogeneous process (31). With regard to the importance of this subject, our study was developed information road map for personalized colorectal cancer 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 aim to develop a personalized screening road map for integrating of the heterologous datasets.

## Methods

The methodology used in this paper is a Mix Method Research (MMR) which combines quantitative and qualitative methodology. This project was conducted in Research institute Gastroenterology and Liver Disease (RIGLD), Shahid Beheshti University in Iran. The comprehensive plan for the screening of colorectal cancer has been designed by RIGLD from the last 18 years (32). This project was performed during 2016 to 2017.

In the first step, a 287-items Checklist 1: (yes T no0) was developed for the assessing required data elements for personalized screening. This checklist contained 4 dimensions sourced from the literature. In this step, all books, articles, research projects, thesis, manual and scientific reports were extracted from MEDLINE, IEEE, Scholar, Web of Sciences, Scopus, ProQuest and databases related to personalized colorectal cancer screening. We synthesized reliable evidence from multiple sources for determining personalized datasets of this checklist. 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 which the evaluating current status of the RIGLD data set by our checklist.

The second step of this paper was a qualitative study. A National information road map of personalized CRC screening was developed in this step. A literature reviews was conducted for identifying components of the information road map. Then both in-depth interview and focus group discussions were performed by clinicians and technical experts of CRC screening. In this study, 37 experts with 15 different fields related to CRC screening were 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 medicines, one anesthesiologist, one surgeon, one psychometric, two social medicines, three nutritionists, two internal medicines 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, screening program) was used. Participants' experiences about components of information road map were collected by interview. We used content analysis and frequency distributions for data analyzing. For final approval of information road map, focus groups with semi-structured discussions were conducted. 18 experts with 8 different fields related to CRC screening & information technology were participated in 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 at least 5 years. Two focus groups were conducted after in-depth interview to approve the final road map.

## Results

This article was presented significant dimensions in the 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 was 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 contains 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 colorectal cancer screening based on personalized datasets were analyzed. The clinical dataset contained 5 items of pathology, surgery, clinical history, colonoscopy, and sign & symptom. Colonoscopy dataset as sub-item of clinical datasets contained 7 subgroups, including patient history for colonoscopy readiness, anesthesia risk assessment, and 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 situational analysis of demographic & a social determinant datasets based on personalized CRC screening. Datasets related to the perspective of the participant was considered as part of the social determinants dataset. However, there were no data elements of perspective of participant in the CRC screening program (Table 3).

Table 1

Situational analysis of Clinical dataset based on personalized CRC screening (59, 60).

Dimensions	Clinical Datasets	Present system n (%)	
Colonoscopy	Items	1/5 (20)	
	Patient history for colonoscopy readiness	Taking anticoagulant drugs, Need for antibiotic prophylaxis, Existence of the pacemaker, Existence of implantable cardioverter-defibrillator, Rectal bleeding	
	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)	

Dimensions	Clinical Datasets	Present system n (%)
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)
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 bowel, chief complaint and duration of its complaint	9/25 (36)

Dimensions	Clinical Datasets	Present system (%)
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 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 (61, 62).

<b>Target Population</b> <b>Patient, Relatives with colorectal cancer, Non- colorectal cancer relatives, Non-patient relatives and Non-patient volunteers</b>		
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)
Life style 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)
Perspective of Participant Dataset	Perceived susceptibility to 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)

Table 3

Situational analysis of genetic dataset based on personalized CRC screening (63, 64).

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)

## Information roadmap for personalized colorectal cancer screening

We developed a national information roadmap for personalized colorectal cancer screening by in-depth interviews and focus group discussion (Table 4). This roadmap was approved by CRC experts. Information leadership is at the top of this roadmap (Fig. 1). This roadmap has 6 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 (33). 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 layer of the roadmap for interoperability of screening dataset are essential factors. Modifiers and descriptors of dataset were defined in fifth layer. Developing of vision, mission and goals of personalized screening as infrastructure of roadmap are dependen to integrated dataset. Information governance and integrated dataset can support continuous editing vision, mission and goals of screening plan by experiences of experts. Also program stewardship and information resources management are an integral parts of the information roadmap of personalized CRC screening.

Table 4

In-depth interview and focus group discussion: Selected comments about 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 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 HI7, 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 personalized screening program is significant.” [Health information Manager]

## Discussions

Personalized datasets of colorectal cancer screening were defined in four dimensions with 275-items. The next step data elements of colorectal cancer screening based on personalized datasets were analyzed. In the final step, the national information road map of personalized CRC screening with 6 layers was developed.

Personalized screening approaches can optimize the efficiency, equity, and safety of cancer screening, but will require precise and comprehensive patient information (34). Recent studies show that accurate and detailed information can support personalized prevention strategies (24, 35, 36). It has been emphasized that screening recommendations require comprehensive patient information (34). Concerning the vital role of patient information in an 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 screening program was entered. While clinical information plays a key role in risk assessment (37). Determining individual risk factors as significant elements of successful implementation of personalizing CRC screening is considerable. Sufficient and up-to-date risk factor's information is an integral part of the more efficient individual screening (22). It is necessary to have a standard documentation tool for evaluation of clinical datasets (38). 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, only 9% of colonoscopy datasets were entered. Patient-centered colonoscopy has become a critical issue in colorectal cancer screening (39). 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 (40). 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 (41). Colonoscopy reporting was poor in this study and data collection process should be revised.

The highest data entry of clinical dimension was related to the pathological dataset in screening program of RIGLD. According to studies, pathology sample is obtained in 30–50% of colonoscopy interventions (42). Accurate histopathological data is a requirement for providing high-quality care services to patients with colorectal cancer (42). 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) (43).

The study proved that the highest data entry of genetic dimension was related to pedigree dataset. Pedigree is useful for tracking and presenting a detailed family history data (44–46). The Pedigree data element is powerful datasets can support genetic dimension in screening approaches (46).

The lifestyle factor plays as an important predictor for screening participant (47–51). While only 11.7% of lifestyle datasets are entered in 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 (52). Therefore, standard datasets for complete and accurate data gathering is one of the requirements of the personalized screening system. In this paper, the standard personalized dataset was developed and the next step present datasets were assessed by this standard dataset and incomplete data elements were identified. Studies show that information gap and deficiencies can affect decision making of clinicians (53–55). Thus, addressing present datasets deficiencies is necessary for better screening decision making.

In addition, 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 (56, 57). To meet this need, we developed a national road map of personalized CRC screening. Data leadership is at the top of this roadmap. Information leadership as a layer of roadmap facilitate effective decision making (33). Successful CRC screening depends on precise and data driven plan (58). In our information roadmap, there is continued interaction between personalized dataset and components of the screening plan such as vision, missions and goals of the program in the development road map. One of the caveats of this study is the fact that the data sample 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 precise screening program is another problem. The developed roadmap identifies various data components of the program and integrates all of information segments. It can be used as a tool for data process reengineering.

## Conclusions

Due to the importance of integrated data in personalized screening approach and the lack of such data sets, researchers were conducted this project. Personalized prevention based on integration dataset plays key role for more efficient implementation of the screening program. Developed road map can be used for integration and interoperability of screening datasets. The efficiency of the decision-making process in screening plan can reduce by data deficiencies. Eliminating data deficiencies can improve the quality of documentation and may lead to improved screening performance. Therefore, the reason of data deficiencies and missing value 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 national roadmap can assist to improve 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 national checklist and a roadmap was developed for 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.

## **Abbreviations**

Mix Method Research (MMR)

Colorectal cancer (CRC)

Research institute Gastroenterology and Liver Disease (RIGLD)

Continuous Quality Improvement (CQI)

TNM Classification of Malignant Tumors (TNM)

Current Procedural Terminology (CPT)

Hereditary Non-polyposis Colorectal Cancer (HNPCC)

Attenuated familial adenomatous polyposis (AFAP)

Familial adenomatous polyposis (FAP)

Immunohistochemistry (IHC)

Adenomatous polyposis coli (APC)

Microsatellite Instability (MSI)

MUTYH (MYH)-associated polyposis (MAP)

Anatomical Therapeutic Chemical (ATC)

## **Declarations**

### **Ethics approval and consent to participate**

The ethical approval was accepted by RIGLD ethical committee.

### **Consent for publication**

Not applicable

### **Availability of data and material**

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

### **Competing interests**

No conflict of interest declared.

### **Funding**

This study was funded with the authors' own contributions.

### **Authors' contributions**

**Elham Maserat:** Design research, gathering data, analysis and Interpretation of data

**Zeinab Mohammadzadeh :** Design research, Analysis and interpretation of data

**Mohammad Reza Zali :** Design research, Analysis and interpretation of data

### **Acknowledgements**

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

## **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. *Nature genetics*. 2019;51(1):76.
2. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1. 0, Cancer incidence and mortality worldwide: IARC CancerBase No. 11. 2013. 2014.
3. Vahdatimanesh Z, Zendehtdel K, Kbari Sari AA, Farhan F, Nahvijou A, Delavari A, et al. Economic burden of colorectal cancer in Iran in 2012. *Medical journal of the Islamic Republic of 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):2564-75.
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. *Journal of the American College of Radiology*. 2019;16(4):580-5.
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 C, Riley C, Huntington M, Foster M, Okuyemi K. Interventions for increasing colorectal cancer screening uptake among African-American men: a systematic review and meta-analysis. 2019.
  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 Expectations*. 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. *Clinical Gastroenterology and Hepatology*. 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. *International journal of cancer*. 2019;145(2):327-37.
  14. Corey L, Valente A, Wade K. Personalized Medicine in Gynecologic Cancer: Fact or Fiction? *Obstetrics and Gynecology Clinics*. 2019;46(1):155-63.
  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 journal of gastroenterology: WJG*. 2014;20(22):6786.
  17. Jain KK. *Textbook of personalized medicine*: Springer; 2009.
  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. 2014 IEEE International Conference on Big Data (Big Data); 2014: IEEE.
  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. *Medical care*. 2008;46(9):S5-S9.
  21. Kuipers EJ, Spaander MC. Personalized screening for colorectal cancer. *Nature reviews Gastroenterology & hepatology*. 2018;15(7):391.
  22. Schilsky RL. Personalized medicine in oncology: the future is now. *Nature reviews Drug discovery*. 2010;9(5):363.
  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. *Applied Immunohistochemistry & Molecular Morphology*. 2019;27(4):270-7.

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. *Journal of digital imaging*. 2016;29(2):160-4.
28. Tan SC. Low penetrance genetic polymorphisms as potential biomarkers for colorectal cancer predisposition. *The journal of gene medicine*. 2018;20(4):e3010.
29. Schaffhausen J. What Precisely Is Precision Medicine? *Trends in pharmacological sciences*. 2017;38(1):1-2.
30. Vidal M-E, 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. Chan JO. Enterprise information systems strategy and planning. *Journal of American Academy of Business*. 2005;6(2):148-53.
32. S RF, Moghimi-Dehkordi B, Maserat E, Zali MR. Prevention strategy of colon cancer at risk population in Iran. *Journal of Medical Council of Iran*. 2010;28(4):444-9.
33. Huvila I. Towards information leadership. *Aslib Journal of Information Management*. 2014;66(6):663-77.
34. Selby K, Bartlett-Esquillant G, Cornuz J. Personalized cancer screening: helping primary care rise to the challenge. *Public health reviews*. 2018;39(1):4.
35. 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. *American Journal of Roentgenology*. 2019:1-7.
36. 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 education and counseling*. 2015;98(10):1280-6.
37. Wang Y, Zhu G, Engel B. Health risk assessment of trihalomethanes in water treatment plants in Jiangsu Province, China. *Ecotoxicology and environmental safety*. 2019;170:346-54.
38. Hoseinpoufard M, Abbasi Dezfouli S, Ayoubian A, Izadi M, Mahjob MP. Hospital compliance with clinical documentation standards: A descriptive study in two Iranian teaching hospitals. *International Journal of Hospital Research*. 2012;1(2):121-5.

39. Sewitch MJ, Azalgara VM, Sing M. Screening indication associated with lower likelihood of minor adverse events in patients undergoing outpatient colonoscopy. *Gastroenterology Nursing*. 2018;41(2):159-64.
40. 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.
41. 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.
42. Haugland HK, Casati B, Dørum LM, Bjugn R. Template reporting matters—a nationwide study on histopathology reporting on colorectal carcinoma resections. *Human pathology*. 2011;42(1):36-40.
43. 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 endoscopy*. 2007;65(6):757-66.
44. Rieger AK, Mansmann UR. A Bayesian scoring rule on clustered event data for familial risk assessment—An example from colorectal cancer screening. *Biometrical Journal*. 2018;60(1):115-27.
45. Li W, Zhao L-Z, Ma D-W, Wang D-Z, Shi L, Wang H-L, et al. Predicting the risk for colorectal cancer with personal characteristics and fecal immunochemical test. *Medicine*. 2018;97(18).
46. 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). *Journal of the American Medical Informatics Association*. 2011;19(1):102-10.
47. Larsen IK, Grotmol T, Almendingen K, Hoff G. Lifestyle characteristics among participants in a Norwegian colorectal cancer screening trial. *European journal of cancer prevention*. 2006;15(1):10-9.
48. 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. *Journal of Medical Screening*. 2000;7(2):99-104.
49. 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.
50. Ioannou GN, Chapko MK, Dominitz JA. Predictors of colorectal cancer screening participation in the United States. *The American journal of gastroenterology*. 2003;98(9):2082-91.
51. 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. *Preventive medicine*. 2004;38(3):276-83.
52. 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. *Radiotherapy and Oncology*. 2014;112(1):59-62.

53. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association*. 2013;20(1):144-51.
54. Waghlikar 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 Summits on Translational Science Proceedings*. 2013;2013:269.
55. Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing electronic note quality using the physician documentation quality instrument (PDQI-9). *Applied clinical informatics*. 2012;3(02):164-74.
56. 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.
57. 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. *Revue medicale suisse*. 2015;11(496):2209-15.
58. 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.
59. Casati B, Bjugn R. Structured electronic template for histopathology reporting on colorectal carcinoma resections: five-year follow-up shows sustainable long-term quality improvement. *Archives of pathology & laboratory medicine*. 2012;136(6):652-6.
60. 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 journal of surgery*. 2015;39(1):259-65.
61. Chen J, He H, Jin H, McAullay D, Williams G, Kelman C, editors. *Identifying risk groups associated with colorectal cancer*. Data Mining; 2006: Springer.
62. 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.
63. Zhou Y, Boardman L, Miller R. Genetic testing for young-onset colorectal cancer: case report and evidence-based clinical guidelines. *Radiology and oncology*. 2010;44(1):57-61.
64. Yuan Y, Zhu L, Xu D, Haixing J, Sun Y, Ding P-R, et al. The prevalence of germline mutations in Chinese colorectal cancer patients with mismatch repair deficiency. *American Society of Clinical Oncology*; 2018.

## Figures

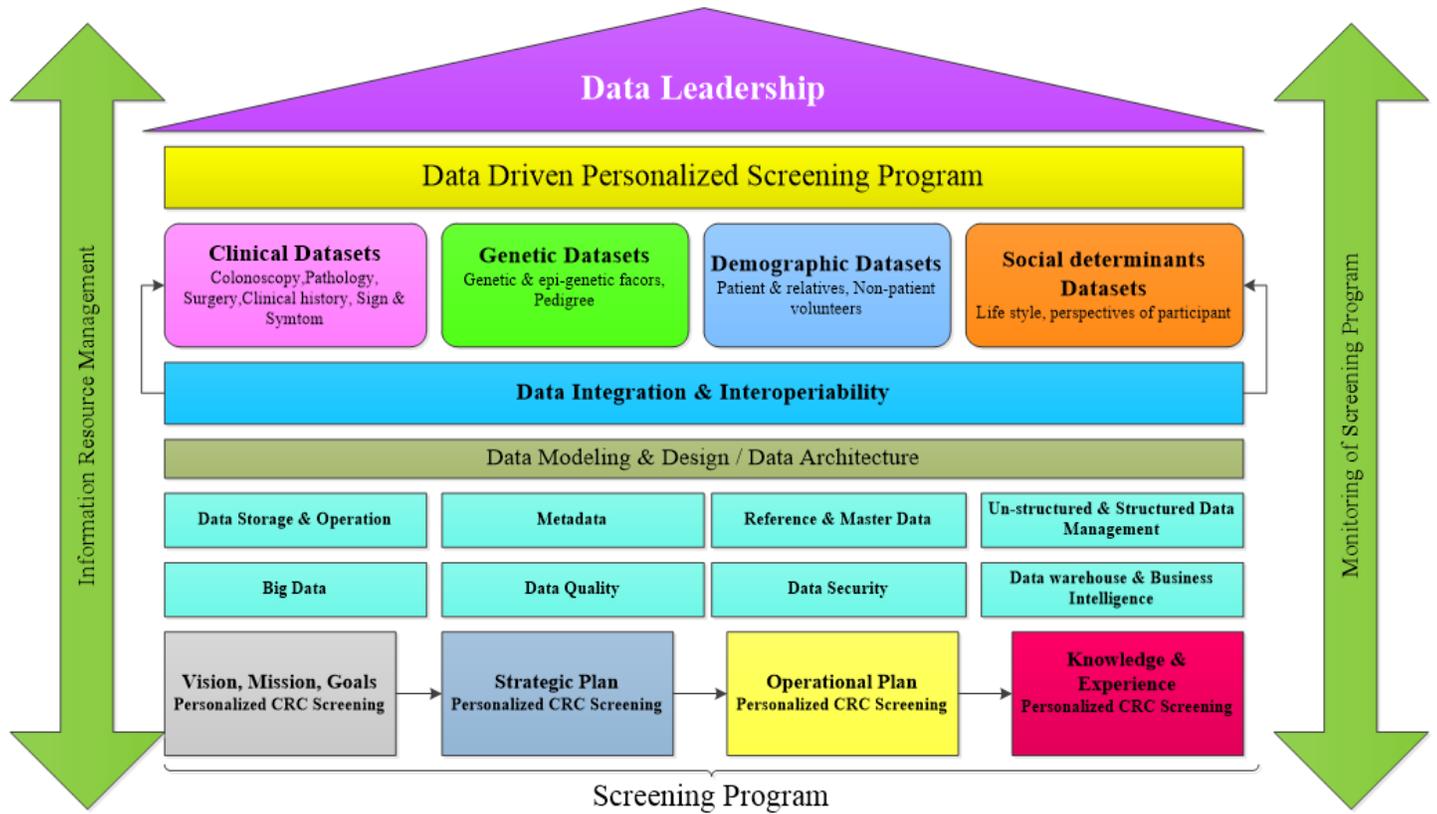


Figure 1

National information roadmap of personalized CRC screening

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Checklist.pdf](#)