Computers in Biology and Medicine 196 (2025) 110755

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A decision support model for colorectal cancer screening

Daniel Corrales [a,b,*], David Ríos Insua [a], Marino J. González [c]

[a] Institute of Mathematical Sciences, ICMAT-CSIC, 28049 Madrid, Spain
[b] Escuela de Doctorado, Universidad Autónoma de Madrid, 28049 Madrid, Spain
[c] Department of Economics, University of La Rioja, 26006 Logroño, Spain

ARTICLE INFO

Keywords:
Colorectal cancer
Decision-making
Multi-criteria decision analysis
Screening
Health policy

ABSTRACT

Background and Objective With minor differences, most national colorectal cancer (CRC) screening programs in Europe consist of one-size-fits-all aged-based strategies. This paper provides a decision analysis-based approach to personalized CRC screening in a general population setting, supporting decisions concerning whether and which screening method to consider and/or whether a colonoscopy should be administered.

Methods We use an influence diagram model characterizing CRC risk with respect to different variables of interest, and including comfort, costs, complications, and information as decision criteria, the last one assessed through information-theoretic measures. The criteria are integrated with a multi-attribute utility model. Optimal screening policies are then computed by maximizing expected utility.

Results The proposed model is used to support personalized individual screening based on relevant features beyond age. It serves to assess existing national age-based screening programs as well as design new risk-based ones. In particular, it suggests replacing current age-based strategies prevalent in many European countries by more personalized strategies based on risk dependent on individual features. Additionally, the model facilitates benchmarking of novel screening devices. Software to implement the model and reproduce the results is included.

Conclusions This work develops a framework for personalized CRC screening that improves upon current age-based screening strategies and highlights how CRC screening strategies could be redesigned and optimized.

1. Introduction

Although colorectal cancer (CRC) is the third most common type of cancer worldwide, making up for about 10% of all cases, only about 14% of susceptible EU citizens participate in CRC screening programs. At the moment, these are mainly *one-size-fits-all* strategies [1] based on age and using fecal testing and colonoscopy (CS). The latter is considered highly invasive, negatively influencing program uptake within the general population [2]. Indeed, [3] reports through the European Health Interview Survey (2018–2020) that the coverage of fecal tests in the population aged 50–74 varies according to the organization of the screening program. In countries with fully rolled-out programs it oscillates from 37.7% in Croatia to 74.9% in Denmark; in turn, in countries without programs or with localized programs it varies from 6.3% in Bulgaria to 34.2% in Latvia. Hence, on the one hand, there is a clear need for accurate, non-invasive, cost-effective screening tests using novel technologies as well as, on the other, for raising awareness about CRC and the importance of its early detection. In addition, genetic, socioeconomic, and behavioral factors can influence the development of CRC and lead to different disease onsets. Recent studies [4] show that early-onset CRC incidence is rising in several countries, pointing at some of the mentioned risk factors as potential causes. Personalized screening strategies that consider these factors seem relevant together with decision tools with guarantees that support experts in their implementation. This is especially important given the discrepancies among the strategies implemented in different countries [5].

As an example, in 2013 the European Union (EU) drafted guidelines for quality assurance in CRC screening and diagnosis, recommending the use of national programs based on fecal immunochemical tests (FIT) and CS [6]. Such guidelines suggested that the identification and invitation of the target population, diagnosis and management of the disease, and the appropriate surveillance of people with detected lesions could be achieved by following and adopting the proposed recommendations. Interestingly, though the screening program structure is similar across many countries, details such as age or test cutoffs differ [7]. In Western Europe, most programs are regional or national and based on FIT or gFOBT (guiac fecal occult blood test), with wide differences in participation rates. In Eastern Europe, countries mostly

* Corresponding author at: Institute of Mathematical Sciences, ICMAT-CSIC, 28049 Madrid, Spain.
E-mail addresses: daniel.corrales@icmat.es (D. Corrales), david.rios@icmat.es (D. Ríos Insua), marino-jose.gonzalez@unirioja.es (M.J. González).

https://doi.org/10.1016/j.compbiomed.2025.110755
Received 24 April 2025; Received in revised form 23 June 2025; Accepted 8 July 2025
Available online 22 July 2025

rely on pilot or opportunistic programs, also based on FIT or gFOBT, with lower participation rates.

Relevant analyses have been developed to assess the importance of various factors influencing the effectiveness of CRC screening programs. A significant portion of the work in the field is dedicated to applying Markov models for cost-effectiveness analyses of screening [8], or to modeling patients' preferences concerning screening methods, taking into account their risk tolerance [9]. An important and complementary approach to further improve the evaluation of program needs, effectiveness, and objectives is the development of CRC predictive models that form the foundation of decision-support tools for screening. In particular, Bayesian Networks (BNs) [10] are used to infer the influence that certain risk factors can have on CRC and how they could affect the decisions made by policymakers. Influence Diagrams (IDs) [11] extend BNs by incorporating decision nodes and utility variables to support multiple criteria decision-making under uncertainty. These tools have proved to be increasingly relevant in medical decision-making [12–14], as they can incorporate and provide information on multiple decision outcomes and objectives.

This paper constructs a decision support model that serves to benchmark CRC screening programs, among other uses. We draw upon earlier work on CRC risk assessment through BNs [15] and complement it to design an ID model that identifies the screening methods adopted, their impacts, and their associated utilities. The model is employed to discuss relevant policy questions concerning CRC screening. For this, Section 2 defines how the problem is structured and present the underlying prediction and preference models. After that, Section 3 describes a set of important use cases including supporting personalized screening decisions; in the light of this, assessing national age-based screening strategies, as well as designing alternative risk-based strategies; and, finally, benchmarking of novel screening technologies that are likely to emerge in the next years.

Software supporting the proposed approach in GeNIe [16] and its Python wrapper PySMILE is available at [17] for reproducibility purposes.

2. Methods

2.1. CRC screening decision problem structure

This section describes the structure of our CRC personalized screening decision support model.

2.1.1. CRC underlying predictive model

Our model stems from a previous BN predictive model [15] developed for CRC risk mapping purposes in a general population setting. The BN aggregates exhaustive expert information and data obtained from a large occupational health assessment study with nearly 2.4M individuals, with information on a selected number of variables and relates modifiable (physical activity (PA), sleep duration (SD), alcohol consumption, smoking status, body mass index (BMI), anxiety, depression) and non-modifiable (sex, age, socio-economic status (SES)) risk factors as well as medical conditions (hypercholesterolemia, hypertension, diabetes) relevant to CRC, assessing such relations through local probability distributions at the nodes. Further details on the variables may be seen in Table 20 in Appendix A. Fig. 1 presents the original BN, with variables in different colors depending on whether they refer to non-modifiable (green) or modifiable (red) risk factors, medical conditions (blue), or the CRC target (purple) variable. Black arrows were initially elicited from expert information, whereas red arrows were discovered and incorporated using our large database.

An extended analysis on the relevance of the selected variables, as well as on procedures to assess its probability tables are detailed in [15], where the BN is used to construct risk maps to segment the population according to risk levels, and detect influential variables in a predictive sense. It is worth noting that, as detailed in [15], all claims from the BN are purely predictive and not causal, as this would require stronger assumptions.

2.1.2. CRC screening decision model

For the construction of the ID supporting CRC screening decisions, the original BN is complemented with a set of variables (chance, decision, and values), together with the corresponding arcs. Before presenting the variables, let us introduce four assumptions used in constructing the model.

1. Rather than the prevalent *one-size-fits-all* strategies [1], essentially based on age, we aim to provide more personalized advice using influential variables whose information does not require more than a general practitioner (GP) checkup. In any case, the model is easily adapted if additional information based on other relevant risk factors or medical conditions is available, as Section 3.1 will illustrate.
2. We adapt to the standard strategy in many EU countries based on applying or not a screening device, say FIT, and, if screened and positive, apply a colonoscopy. The objective of the decision model is to suggest to a person with certain features the most convenient screening policy.
3. The decision made should be based on the selected variables, the information provided by the screening results, the costs, the entailed complications, and the patient's comfort. In doing this, we focus on the short-term outcomes of the screening intervention, and refrain from considering pointers to longer-term outcomes like, e.g., expected QALYs gained. This is motivated mostly by our interest in benchmarking novel screening devices, as sketched in Section 3.4, for which little information will be typically available. The model allows for assessing the relative importance of various criteria as described below through their integration into a multi-attribute utility function.
4. Suggested screening strategies will be chosen as maximum expected utility (EU) alternatives [18], with the utility vision of either the patient, the doctor, or the health policy maker, as later discussed.

These hypotheses, specially the third one focusing on short-term outcomes, are broadly aligned with the recommendations from the European Network for Health Technology Assessment [19]. Based on them, we describe the remaining elements required to structure the problem qualitatively.

New nodes and arcs

To complete the ID design, the following nodes and arcs are added to the BN in Fig. 1

Decision nodes. The *screening method* implemented is considered as a decision variable. Its potential alternatives are the currently most common CRC screening methods, to wit: *gFOBT*, *FIT*, *blood-test*, *stool DNA (SDNA) test*, *computed tomography colonography* (CTC) and *colon capsule* (CC). Obviously, we also include the possibility of conducting *no screening*.

A second decision node refers to the possibility of *performing a colonoscopy*. It is a successor of the previous one and, hence, we consider the possibility of performing it after (or without) screening. Note that it is standard in many European countries to perform a colonoscopy if screening suggests the presence of CRC, say through a positive FIT. On the other hand, several countries opt for directly performing a colonoscopy on susceptible patients. Both possibilities are thus covered within our model and, even enriched, as they can be combined, and further, more individualized information, beyond age, is used to support the corresponding screening decisions.

Chance nodes. Besides the chance nodes from the initial BN, we include three additional chance ones. First, we consider the potential *complications* associated with the eventual screening and colonoscopy interventions [5], which include *bleeding*, *retention*, *perforation*, *death* and, obviously, *no complications*. The other two nodes refer to the results of both interventions, *screening results* and *colonoscopy results*,

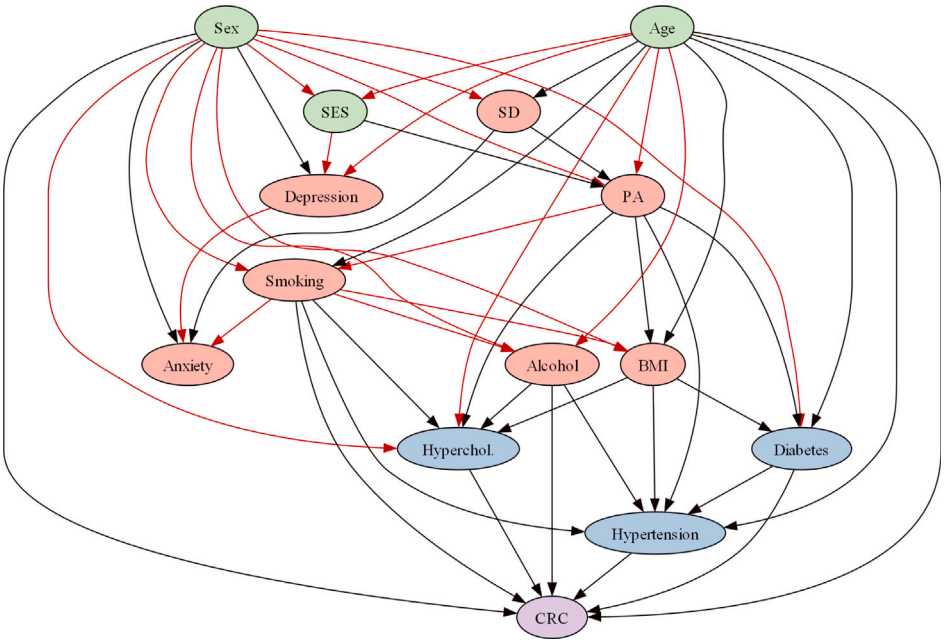

Fig. 1. Originating CRC Bayesian network [15].

assimilated to two possible reports: *predicted true* (interpreted as screening or colonoscopy suggesting the presence of CRC) and *predicted false* (interpreted, otherwise). We include as well a *No result* state to handle the case when the corresponding intervention is not actually performed.

Value nodes. We introduce a multiple criteria preference model [18] to support CRC screening decisions. As primary criteria, the model includes the *cost of complications*, the *cost of the intervention*, its *comfort* and the *information provided by screening and/or colonoscopy*. Finally, we include a value node aggregating the four criteria through a multi-attribute utility function, later specified.

Arcs from the original BN are preserved. Besides, we include arcs which essentially reflect our starting hypothesis and the information relations between decisions and values.

First, according to hypothesis 1, we initially assume that the screening decision is made knowing variables that are easy to obtain or request (BMI, Age, PA, Sex, Alcohol consumption status, Smoking status, Sleeping duration (SD)), hence the arcs from such variables into the screening node. Notwithstanding these, the diagram is easily modifiable if we know other variables from the patient, like its eventual hypertension, as Section 3 illustrates.

The arc between the decision nodes and the arc connecting the screening results with the colonoscopy decision reflect hypothesis 2 and permit covering the standard screening protocols within the EU. In turn, the arcs going into the value nodes reflect hypotheses 3 and 4. Note that some arcs are displayed in a lighter color. This is because the value function defined in Section 2.2.2 will depend on the parent nodes of CRC and the decision nodes. However, to simplify the analysis and visual structure of the model, we just kept the original color of the arcs going to screening and colonoscopy results, and CRC.

2.1.3. Final influence diagram

Fig. 2 reflects the produce ID structure. In line with our comments, the possibly influential variables considered for the screening decision include those whose retrieval does not require further than a GP checkup. However, the socioeconomic situation together with the presence of depression and anxiety have not been taken into consideration as optimizing expected utility based on these variables generates obvious ethical conflicts. These, however, could perfectly be used in future extensions of the model, e.g. for addressing screening uptake in the population. The diagram was presented to several medical experts who validated it for concept and meaning.

2.2. Quantifying the CRC screening decision model

This section describes the probability and preference models included in the proposed ID to facilitate CRC screening decision support.

2.2.1. Probability models at chance nodes

The ID inherits the probability models at nodes from the original BN. We refer to [15] for details concerning how the probability tables were built based on prior distributions and our available occupational health assessment database. Note that, as a consequence, the probabilities available were assessed with some uncertainty through posterior predictive Dirichlet distributions. We use here just the mode of such distributions.

For the other three chance nodes, we used public sources, mainly [5], to obtain the required parameters. In particular, Table 21 (Appendix B) includes the sensitivities and specificities of screening methods and those of colonoscopy, from which we deduce the required probabilities in the *results of screening* and *results of colonoscopy* nodes, given the presence (or not) of CRC and the method chosen. As an example, the probability table at node *result of colonoscopy*, which depends on the actual presence of CRC and the use of colonoscopy would be as in Table 1, with e.g. the probability of reporting *Does not have CRC*, given that the person goes through a colonoscopy and she actually has CRC, being 0.03.

Similarly, the probabilities in the first seven columns of Table 22 (Appendix B) were used to build the probability table of *complications* associated with each screening method and colonoscopy. Full details of the tables are available in the accompanying software [17].

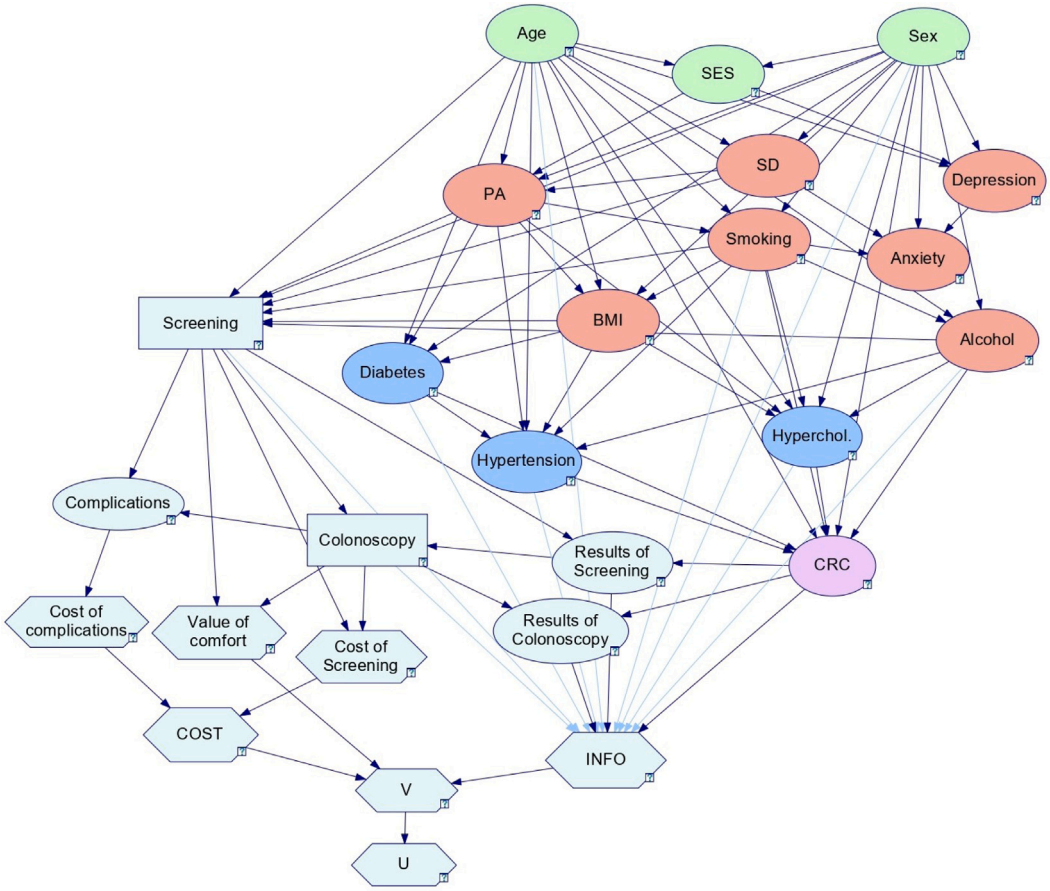

Fig. 2. CRC screening decisions influence diagram.

Table 1
Probabilities of colonoscopy results, given antecessors.

Colonoscopy	Yes		No	
CRC	Yes CRC	No CRC	Yes CRC	No CRC
Has CRC	0.97	0.01	0	0
Does not have CRC	0.03	0.99	0	0
No result	0	0	1	1

2.2.2. Preference models

Ley us discuss first the preference model for each criterion and then how to aggregate them through a multi-attribute utility model.

Single criterion preferences

Cost c_{int} of intervention. This criterion aggregates the costs of the eventual screening and colonoscopy interventions. Both should be minimized. We adopt the costs in € for France collected in [5] for France, displayed in Table 21 (bottom row).

Cost c_{comp} of complications. These are assessed through their entailed expected costs in € for France obtained from public sources in [5], available in Table 22 (right column).

Comfort com**.** Lacking a natural attribute to assess intervention comfort, we used a *constructed attribute* [20] with four decreasing levels, the best level (4) referring to *no screening*; the worst one (1), corresponding to *colonoscopy*. Though, in principle, we could assess this in a personalized manner, we constructed the scale reflected in Table 2, later validated by several experts and patients.

Torrance et al. [21] provide scales within their HUI:2 quality of life system covering several criteria with one of them, pain, close to ours. However, their effort is addressed towards assessing quality of life over the years by creating a utility function, whereas ours, in line with our short-term health assessment focus [19], is more geared towards avoiding uncomfortable episodes.

Importantly, when both screening and colonoscopy are implemented in a patient, the comfort value would be that of colonoscopy, that is 1, since this is perceived as much more uncomfortable than any of the six analyzed screening methods.

Information v_{info} provided. The more complex to understand and elicit attribute refers to short-term informational effects of interventions. These essentially entail moving from a state of uncertainty based on the probability of a person having CRC (and that of not having it) given their features and the same probability (and its complementary) when we know, as well, the screening and/or colonoscopy results.

As an example, consider the case of a male adult, age 44–54, with normal sleep duration, physically active, normal weight, non-smoker, and with low alcohol consumption, who is negative in all medical conditions considered. We shall use it as a *benchmark* patient below. Table 3 displays the CRC probability for this individual and his eventual probabilities after screening with FIT and colonoscopy for various results. For instance, a positive FIT would move the prior uncertainty $(0.0009, 0.9991)$ to the posterior $(0.02, 0.98)$. We would like to assess the information provided by various results taking into account such probabilities (and the complementary ones of not having CRC). Therefore, we seek ways to evaluate such changes in uncertainty, that is, to assess the value of the information that screening provides concerning CRC presence. This will be based on three terms well-known to clinicians and policymakers: the CRC probability, and the sensitivity and specificity of screening tests. The rationale follows the argument from several works [22–24] suggesting the use of the *mutual information* as a measure of diagnostic test performance. That said, the value of information v_{info} provided by a screening strategy concerning the presence of CRC will be given by a quantity that we term *normalized* or

Table 2
Comfort levels for interventions.

com	Description	Interventions
4	The patient does not experience any discomfort	No screening
3	The patient experiences a minor discomfort or the test implies a small inconvenience: time lost, emotional difficulty, or minor physical pain.	FIT, gFOBT, sDNA, Blood-test
2	The discomfort experienced by the patient is noticeable. There is a noteworthy emotional aversion and a few moments of physical discomfort.	CTC, CC
1	The discomfort is very significant. The test causes some periods of pain resulting in remarkable distress.	Colonoscopy

Table 3
Probabilities of CRC depending on screening outcomes for benchmark patient.

$p(CRC)$	$p(CRC\|FIT-)$	$p(CRC\|FIT+)$	$p(CRC\|FIT+,COL-)$	$p(CRC\|FIT+,COL+)$
0.0009	0.0002	0.02	0.0006	0.65

Table 4
No screening v_{info} for benchmark.

Scr	No screening					
R_s	No pred					
CRC	False			True		
Col	No Col	Colonoscopy		No Col	Colonoscopy	
R_c	No pred	Pred false	Pred true	No pred	Pred false	Pred true
v_{info}	0.0	0.12	−11.44	0.0	−509.19	654.93

relative pointwise mutual information (RPMI) and that is defined through

$$v_{info}(crc, r_s, r_c) = \left[\log\left(\frac{p(crc|r_s)}{p(crc)}\right) + \log\left(\frac{p(crc|r_s, r_c)}{p(crc|r_s)}\right)\right] \Big/ H(CRC). \tag{1}$$

Intuitively, the first log term accounts for the change in uncertainty on the state of CRC after performing screening, while the second log term accounts for the change in the uncertainty of CRC when performing a colonoscopy after screening. Besides, $H(CRC)$ refers to the entropy of CRC. Appendix D provides a detailed construction and explanation of this value function.

Importantly, this proposal facilitates acknowledging some difficulties usually encountered in the preference elicitation scenarios considered in the CRC domain. For instance, in relation to the psychological cost of unnecessary assessments associated with false positives [25] and the possibility of a delay in the CRC diagnosis and subsequent dissuasion of participants for later assessment [7], the proposed mechanism weighs these situations through their corresponding uncertainty.

As examples, Tables 4 and 5 respectively provide the v_{info} for our benchmark patient when not undertaking screening and when performing FIT. Positive values correspond to correct predictions; negative values, to wrong ones; and, finally, zero indicates no change in uncertainty. Values corresponding to cases in which CRC is present are larger in magnitude as they have a low probability. This conforms with the argument that a missed CRC positive scenario is much worse than a misdiagnosed healthy patient [26]. When screening is performed, not much information is gained if the result predicts *false*, the most likely scenario in general. However, the information provided is much larger when the prediction is *true*, further increasing its value when a posterior colonoscopy is performed.

Fig. 3 plots the v_{info} function for all possible values of the probability of having CRC facilitating comparison of screening methods in terms of information. Observe that blood-test, CTC, and gFOBT are bounded above by the rest of screening methods in terms of the information measure used. From this, we conclude that we could discard blood-test and CTC as they are equal or worse to FIT in the four criteria *(information, comfort, specificity, sensitivity)* considered. This is not the case, however, for gFOBT as it is the cheapest method and has non-dominated performance metrics.

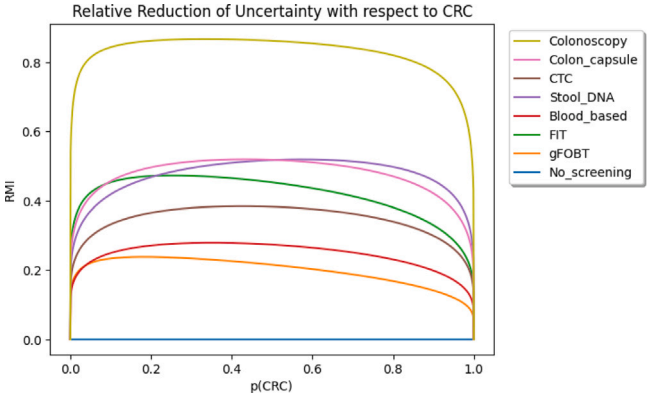

Fig. 3. v_{info} for various CRC screening alternatives.

Further notice how in a low prevalence scenario highly specific tests, like FIT, provide more information, whereas highly sensitive tests, like sDNA, do so when the probability of CRC is larger. That difference in sensitivity and specificity is also what provokes the asymmetries in the curves in Fig. 3.

Multiple criteria preference aggregation

As the final step for constructing our decision model, let us aggregate the criteria through a utility function u, first employing a multicriteria value function, then transforming it to take into account risk aversion, see [27] for a detailed conceptual description.

We first aggregate the costs in € so that $cost = c_{int} + c_{comp}$ for each intervention. The information provided by the screening strategy is already in compact form, and, thus, we next define the global value function as a weighted aggregation of the intervention costs and information, taking into account comfort. To wit, for a fixed comfort level k, under reasonably general conditions [27], we use a general weighted additive value model. However, after initial numerical experiments with the original information and cost scales, we decided to $\log_{10}$ this last one, finally adopting the model

$$v(cost, info, comf = k) = \lambda_k \times v_{info} - \log_{10}(cost + 1),$$

Table 5
FIT v_{info} for benchmark.

Scr	FIT											
R_s	Pred false						Pred true					
CRC	False			True			False			True		
Col	No Col	Colonoscopy		No Col	Colonoscopy		No Col	Colonoscopy		No Col	Colonoscopy	
R_c	No pred	Pred false	Pred true	No pred	Pred false	Pred true	No pred	Pred false	Pred true	No pred	Pred false	Pred true
v_{info}	0.09	0.12	−2.97	−196.80	−706.08	466.52	−2.59	0.04	−151.00	448.05	−58.63	966.01

Table 6
Expected information of different policies for reference patient.

Screening	No screening		FIT				sDNA			
Result of Scr	No pred		Predicted false		Predicted true		Predicted false		Predicted true	
Colonoscopy	No Col	Col	No Col	Col	No Col	Col	No Col	Col	No Col	Col
Exp. v_{info}	0.0	0.532	0.049	0.187	5.722	15.802	0.086	0.134	0.911	4.394

Table 7
Expected information of interventions benchmark.

	No scr.	gFOBT	FIT	Blood	sDNA	CTC	CC	Colonos.
v_{info}	0	0.129	0.245	0.121	0.197	0.159	0.225	0.532

where λ_k, textcolorblue designated comfort parameter, is a weighting factor that depends on the comfort of the screening strategy and serves as a trade-off between information and log-cost.

The elicitation of the comfort parameters is delicate. We adopted the following strategy. Assume that only one test is performed at each time and the amount of information corresponds to the reduction of uncertainty through a single independent test, that is, $v_{info} = MI(CRC, R_s)/H(CRC)$. Consider two screening methods with the same comfort level k, yet different costs and values, say $(info_1, cost_1)$ and $(info_2, cost_2)$, where, typically, the more informative the method is, the more expensive it will be. Assume that no method dominates the other, in the sense of being cheaper and more informative, and that the individual declaring his preferences reveals that he favors $(info_1, cost_1)$ to $(info_2, cost_2)$, represented as $(info_1, cost_1) \succ (info_2, cost_2)$. Then, we interactively ask the respondent for a $\overline{cost}$ value (smaller than $cost_2$) such that $(info_1, cost_1) \sim (info_2, \overline{cost})$. These options would have the same value, that is,

$$\lambda_k \times info_1 - \log_{10}(cost_1 + 1) = \lambda_k \times info_2 - \log_{10}(\overline{cost} + 1), \tag{2}$$

and we solve for $\lambda_k = (\log_{10}((cost_1 + 1)/(\overline{cost} + 1))/(info_1 - info_2))$. Such value would then be subject to standard consistency checks, see e.g. [28]. In particular, note that we would expect that for lower discomfort, the value of λ should be larger, that is, $\lambda_4 > \lambda_3 > \lambda_2 > \lambda_1$, so that information is perceived as more valuable when obtained from a more comfortable screening tool.

A more robust estimation of the parameter would perform this exercise for each pair of screening methods at each comfort level, obtain the corresponding estimations, and reconcile them through their median. This final estimation would typically be more robust than just a single estimation made out of a chosen pair of screening methods, but has the drawback of requiring a larger elicitation effort on behalf of the decision respondent. Appendix C provides a full elicitation exercise for the following example.

Example 1. For reference purposes, consider the benchmark patient to elicit the comfort parameters. Table 7 presents the v_{info} provided by all interventions. Fig. 4 contains the cost and information provided by the methods for the benchmark patient as well as their comfort. Note that, in this case, the screening alternatives that are non-dominated are *no screening*, *gFOBT*, *FIT*, and *colonoscopy*. This will not always be the case as our information value assesses uncertainty and depends on the probability of having CRC, see Section 3.5.

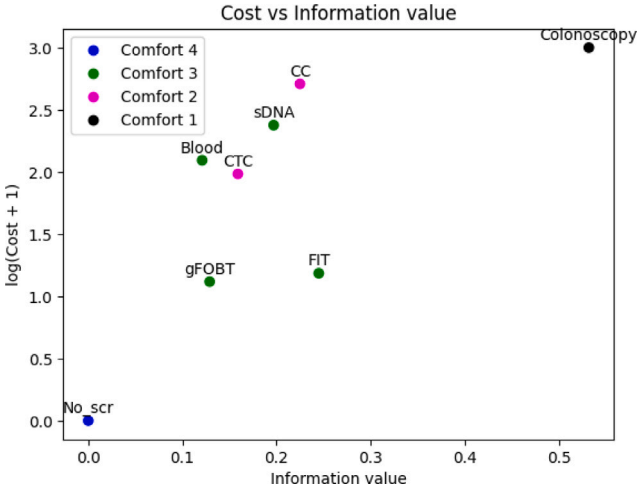

Fig. 4. Scatterplot.

To illustrate the elicitation process, consider comfort level 2. CTC and CC respectively provide a v_{info} of 0.159 and 0.225, their aggregated costs being 96.51€ and 510.64€. Suppose the individual declares preferring CTC to CC. He is interactively asked how much should the CC cost be so that they are indifferent to one or the other. Assuming that the interview converges to a cost of 180€, using (2) we obtain $\lambda_2 = 4.17$.

For comfort level 3, FIT and gFOBT respectively provide a v_{info} of 0.245 and 0.129, while their aggregated costs are 14.34€ and 12.14€. Suppose the patient declares preferring FIT and that the indifference cost for gFOBT is 3€. Then, we obtain $\lambda_3 = 5.04$. Repeating this process for all pairs of methods at this comfort level and taking their median, we obtain $\lambda_3 = 6.80$. Table 23 (Appendix C) reflects the full elicitation exercise.

Comfort levels 1 and 4 are slightly different as we only have one option for them. For comfort level 1, let us consider a synthetic test, providing a v_{info} of 0.4 and an indifference-associated cost with respect to colonoscopy of 300. In that case, $\lambda_1 = 4.01$. For comfort level 4, both the v_{info} and cost are 0. Therefore, λ_4 has no impact in the calculation of the value function, and we just assign λ_4 so that the monotonicity of the λ's is preserved, leading to Table 8. △

Table 8
Values of λ parameters.

Parameter	Value
λ_1	4.01
λ_2	4.17
λ_3	6.80
λ_4	7

Table 9
Utility parameters.

Param.	Value
a	1.015
b	0.872
ρ	0.039

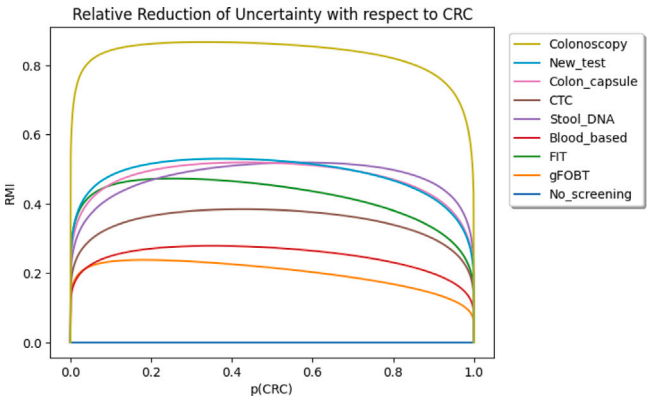

Fig. 5. Plot of v_{info} at different $p(CRC)$ levels.

Once the value function is elicited, assuming (constant absolute) risk aversion [27] we adopt the following expression for the utility function

$$u(cost, info, comf) = a - b \times \exp(-\rho v(cost, info, comf)),$$

where ρ is the risk aversion coefficient and a and b are scaling constants constraining the utility to the $[0, 1]$ interval for the three reference alternatives, determined using, e.g., the classic probability equivalent (PE) method [29].

Example 2 (*Cont.*). In our problem, we assume that the presence of CRC is uncertain and, hence, the best outcome in terms of information collected will be detecting a high-risk patient with a very specific test and then performing a colonoscopy. In turn, in terms of information, the worst outcome would be not doing anything as it entails no added value. Concerning cost, the best option would be a costless test, while the worst option would be a very expensive test with all complications, producing an additional cost to the patient. Thus, we choose as reference the pairs ($cost^* = 0$€, $info^* = 15.75$) and ($cost_* = 8131.71$€, $info_* = 0$).

Suppose that the reference to assess the risk aversion coefficient is ($cost = 50$€, $info = 4.1$).[1] Assume all interventions have comfort level 3 and we obtain through the interview that the PE is 0.7. We then consider the system

$$\begin{cases} a - b \times \exp(-\rho v(8131.71, 0, 3)) = 0 \\ a - b \times \exp(-\rho v(0, 15.75, 3)) = 1, \\ a - b \times \exp(-\rho v(50, 4.1, 3)) = 0.7 \end{cases}$$

leading to Table 9. △

We employ such parameters in our use cases in Section 3.

3. Results

Once with our ID built and validated, we illustrate its application in relevant use cases in relation to providing personalized screening advice; assessing national age-based screening strategies; designing risk-based national screening strategies; and, benchmarking novel screening technologies.

3.1. Personalized screening advice

Let us exemplify a few cases of individuals for which our model would propose different screening advice showcasing how our approach personalizes screening recommendations, as summarized in Table 10.

[1] These values have been chosen to maximize the tool's capability for classification as a validation mechanism, maximizing the F1-score. See end of Section 3.3.

Start with the benchmark patient from Section 2.2.2 whose probability of having CRC was 0.00085. Running our model, the decision with maximum EU for him would be not performing screening (EU 0.143). As a comparison, the decision with the second highest EU (0.142) is FIT (followed by a colonoscopy if the predicted result is positive).

Suppose now that the preference model is more risk-seeking, reducing the risk aversion coefficient ρ from 0.039 to 0.005. Then, the optimal suggested policy would be FIT (EU 0.147), followed by a colonoscopy if the FIT result is positive (and no colonoscopy if negative). Furthermore, all of the tests, except for CC and blood-test, are preferred to no screening. Intuitively, in this problem, a risk-seeker would be willing to make a larger expense for the information.

As mentioned, the model is easily adaptable when additional medical information is available, beyond the seven variables obtainable through a GP visit. This is done by just adding the corresponding arrows that go from the new variables of interest to the decision node. As an example, suppose that we also know that the benchmark is hypertense and has diabetes. His probability of having CRC is much larger (0.0039). Using the model, the advice would be to screen with sDNA (EU 0.146), followed by a colonoscopy if the prediction is positive.

Suppose now that for a given person we have access to an exogenous variable that changes the probability of CRC to 0.1, e.g. based on knowing that the patient has CRC family antecedents. In this situation, the recommendation for the first decision would be FIT with an EU of 0.183; the policy with the highest EU would be FIT followed by no colonoscopy even if the result is positive. The reason behind this is that the change in uncertainty that a test can produce at these relatively high levels of CRC probability is not worth the costs associated with a colonoscopy. Indeed, the information provided by a FIT in this case is already quite significant as it changes the prior probability of CRC from $p(CRC) = 0.1$ to $p(CRC|FIT+) = 0.710$ when the result of FIT is positive, rendering somewhat redundant the results of a colonoscopy.

3.2. Assessing the French national screening strategy

Given the previous example, our goal now is to assess, the current *one-size-fits-all age-based* French screening strategy. This case is chosen because of data availability [5]. Note though that the methodology is general and may be easily replicated in many other countries with similar strategies.

More precisely, the target population for CRC screening are citizens in the age range [50–74] which do not have any hereditary condition or familiar CRC antecedents; when FIT is positive, participants are suggested to undertake a colonoscopy [30]. Let us present three cases that are not properly prioritized in the current strategy.

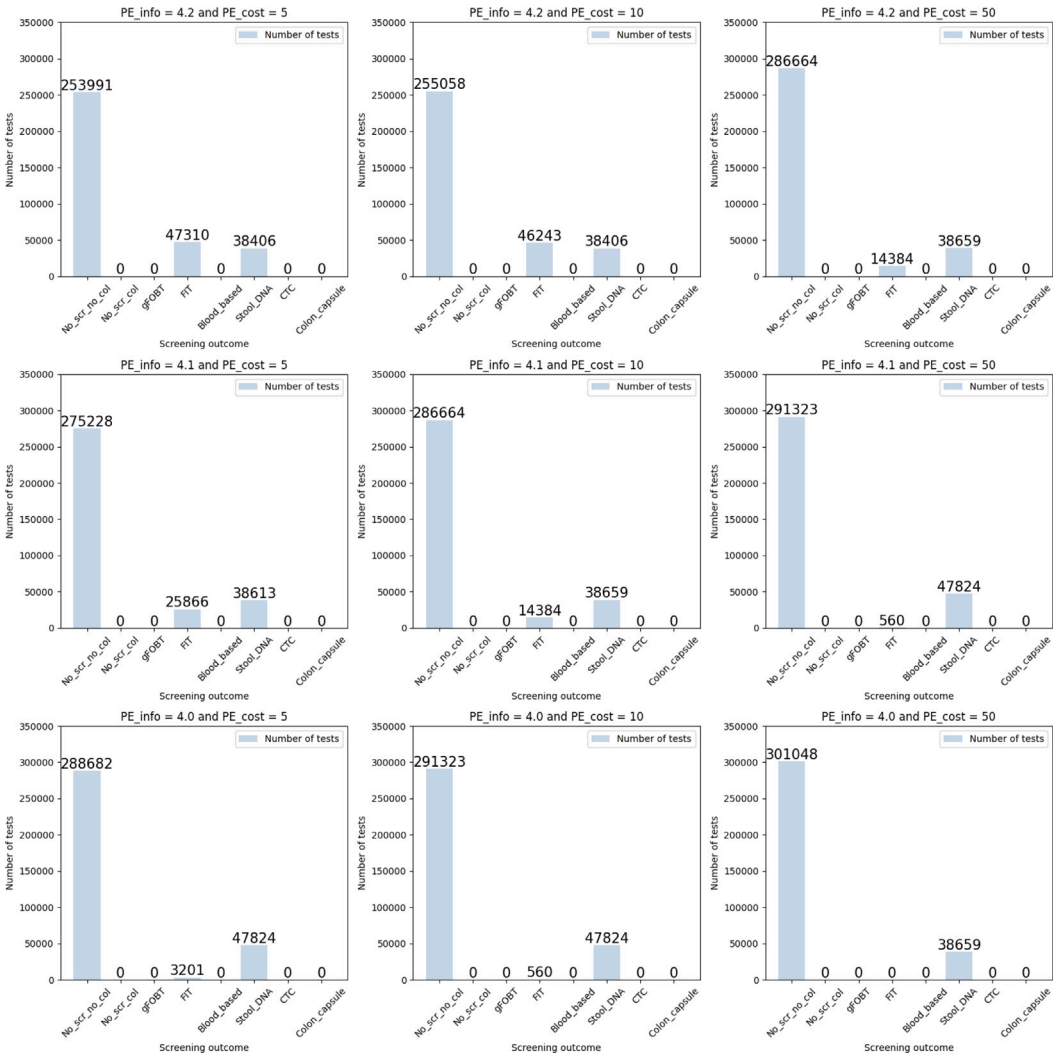

Fig. 6. Plot with different PE info and PE cost.

Table 10
Personalized screening strategies in the four cases.

	1st recom	EU	2nd recom	EU	EU Pred false		EU pred true	
					No col	Col	No col	Col
Benchmark patient	No scr	.143	FIT	.142	–	–	–	–
BP risk seeking	FIT	.147	sDNA	.145	**.139**	.133	.293	**.370**
BP added ev.	sDNA	.146	FIT	.145	**.084**	.056	.289	**.536**
Exogen. var. $p = .1$	FIT	.183	sDNA	.173	**.131**	.098	**.631**	.554

1. Consider a man with age [54–64], normal sleep, normal BMI, physically active, non-smoker, low alcohol consumption and not having diabetes or hypertension. His probability of having CRC is 0.0022. Because of his age, he would be asked to undertake FIT. However, the proposed model suggests that it is better to administer sDNA (EU 0.145) rather than FIT (EU 0.144). Both would be followed by a colonoscopy if the screening prediction is CRC-positive.
2. Consider now a man with age [44–54], with similar characteristics but having diabetes and hypertension. His probability of having CRC is 0.0039. Because of his age, he would not be called to participate in the screening program. However, the model suggests this patient should undertake sDNA (EU 0.146), followed by colonoscopy if sDNA is positive.
3. Finally, for the case of a man with age [44–54], overweight, normal sleep, high-alcohol consumption, physically inactive, and ex-smoker, the probability of having CRC will be 0.0018. The proposed model suggests that the optimal screening method is FIT with 0.143, followed by a colonoscopy if its result is positive.

These examples suggest that the current age-based strategy may fail in various ways. As an example, the second patient should be prioritized over the first one within a screening program. Further, sDNA may be more recommendable for higher-risk patients, while FIT is a better choice for moderate-risk patients. Predictive models for decision-making can help redistribute resources to produce more efficient screening programs, as we show next.

3.3. Designing a national personalized screening strategy

Let us discuss now how the design of a national screening strategy could be based on our model. For this, we use a database with around

Table 11
Comparison of strategies with no constraints (top row) and with constraints (third row). Operational limits are shown in the second row.

	Nothing	Colon.	gFOBT	FIT	Blood	sDNA	CTC	CC	Total
No lim recom.	291 323	0	0	560	0	47 824	0	0	48 384
Oper. limit	∞	3000	30 000	42 000	7000	6000	2000	2000	–
Final recom.	291 707	0	0	42 000	0	6000	0	0	48 000
National	290 633	0	0	49 074	0	0	0	0	49 074

350 000 individuals whose records were kept for testing purposes for the BN model in Section 2.1.1, as [15] fully describes. We use all thirteen available variables in the network for this use case. As our database does not include screening or colonoscopy results, we simulate them based on the sensitivity and specificity information of various interventions available in Table 21.

When designing a screening program, it is important to take into account resource constraints on e.g. the maximum number of colonoscopies and screening operations performable, because of lab, personnel, and device availability, as well as due to budget limits. The issue is, then, how do we best allocate such resources using the decision support model in Section 2. Our approach will assign screening methods in order of maximum EU: individuals with higher EU will be offered screening earlier as they are assimilated with the population benefiting more from screening. The screening choice will be that providing the highest EU. Once the n_i tests available for the ith screening method are saturated, we remove it from the list of available methods and search for those with the maximum EU among the remaining ones. The process continues until we reach all available test limits or cover the entire targeted population. We assume that there is no limit on the available colonoscopies after a positive screening prediction as these are fundamental for a correct diagnosis, a sound assumption as it corresponds to the standard health practice. On the other hand, primary colonoscopies (those directly delivered without screening) will indeed be limited.

Let us see how this strategy performs by comparing three setups:

1. We assume there are no constraints when applying the new strategy.
2. Constraints are set on the number of tests for each screening method, as expressed in Table 11 middle row. Specifically, we limit the maximum number of primary colonoscopies to 3000; gFOBT to 30 000; FIT to 42 000; blood-based tests to 7000; sDNA to 6000; CTC to 2000; and, CC to 2000, therefore being able to cover 92 000 individuals (out of the 350 000).
3. Finally, the current national age-based strategy.

With the parameters developed throughout the text, the distribution of recommended tests and results would be as Table 11 shows, where the first row indicates tests administered when no constraints are included; the middle row indicates test constraints; the third row indicates the number of tests when constraints are included; and, finally, the results attained with the current national strategy. Interestingly, only sDNA and FIT are recommended as screening technologies. The excess of recommended sDNA tests is distributed by administering FIT and implementing no screening. No direct colonoscopy would be recommended in this general context. For 291 707 patients, no screening is preferred to the remaining alternative screening methods even when setting operational limits.

We apply the three strategies described and obtain results for 200 simulations. Tables 12–14 contain their confusion matrices for CRC classification with values corresponding to the mean over 200 simulations and their standard deviations. Table 12 shows how the current age-based strategy has a poor performance in detecting patients with CRC, i.e. a poor sensitivity of 0.36. In turn, the proposed unconstrained new strategy, Table 13, increases detection sensitivity to 0.45; however, this comes with the trade-off of a decrease in precision (0.82 to 0.64), resulting in a larger number of false positives. Concerning cost, the unconstrained new strategy entails an increase in the average cost

Table 12
Mean classification results with current strategy. Cost per patient: 7.24€.

	Predicted no CRC	Predicted CRC
No CRC	339 472.1 ± 4.0	16.9 ± 4.0
CRC	139.8 ± 4.4	78.3 ± 4.4

Table 13
Mean classification results with new strategy, no constraints. Cost per patient: 44.62€.

	Predicted no CRC	Predicted CRC
No CRC	339 425.2 ± 8.2	63.8 ± 8.2
CRC	121.3 ± 3.1	96.7 ± 3.1

Table 14
Mean classification results for new strategy with operational constraints. Cost per patient: 12.79€.

	Predicted no CRC	Predicted CRC
No CRC	339 466.5 ± 4.9	22.5 ± 4.9
CRC	134.8 ± 4.1	83.2 ± 4.1

(44.62 € vs. 7.24€ per patient). To account for that cost increase and the precision reduction, we set operational limits (second row of Table 11) on the number of tests usable for each method. Table 14 shows the corresponding classification metrics from this strategy, with more moderate costs (12.79 € vs. 7.24€ per patient). There is still an increase in sensitivity (0.37) and a decrease in precision (0.79), but both are more subtle. Depending on the cost-information trade-off and risk attitude, the take on the results of the new strategy is that such investment may be recommended to reach higher sensitivity levels and thus increase the number of detected patients in screening programs. However, this is not always manageable and demands an operational limit in more realistic scenarios.

Table 15 summarizes the proposed risk-based screening strategy, and compares it with the current national screening strategy in Spain, in terms of its design. As an example, a man under 50 with high risk (due to e.g. high alcohol consumption, overweight, ex-smoking) would be identified by the model as a higher risk patient than a healthy man above 60.

As a final validation, assume that the current system can cover 49,074 FIT tests, the number of people in the age range [54,64] in our 2016 dataset. Assuming this, we compare the current strategy with our model by taking the 49,074 patients with highest EU for FIT and performing 200 simulations to assess performance and cost differences. Table 16 shows a subtle increase in sensitivity when using the model, correctly detecting an average of one more patient (an increase of 1%) and reducing the cost of the whole strategy by around 33 000€. Although the improvement might seem relatively low, when extrapolated to a real-sized population, the method can save many lives and money. For example, in France and Spain, respectively with populations of $8,591,286$ and $6,583,183$ within such age range, the implementation of the model would detect about 175 and 134 more positive patients in both countries.

3.4. Benchmarking of new screening devices

Given the increasing importance of CRC from a public health policy perspective, it is likely that, in the near future, there will be novel CRC

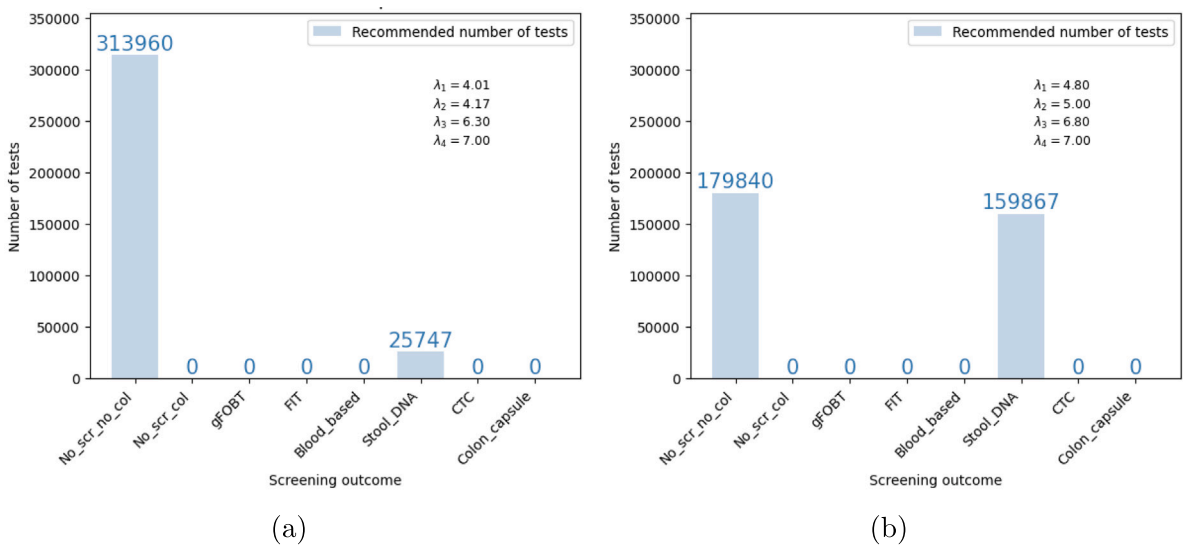

Fig. 7. Screening count distribution for different λ values.

Table 15
Comparison of current Spanish screening strategy vs. proposed risk-based approach. θ_1 and θ_2 depend on resources available.

Current Spanish strategy	Proposed risk-based strategy
• If patient >50 years old: Send FIT invitation. • Else: No screening, invitation.	For patient with features x: • if $p(CRC\|x) < \theta_1 \rightarrow$, no screening invitation. • if $\theta_1 \leq p(CRC\|x) < \theta_2 \rightarrow$, FIT invitation. • if $p(CRC\|x) \geq \theta_2 \rightarrow$, sDNA invitation.

Table 16
Mean classification results for new strategy on 49 074 patients with highest FIT utility. Cost per patient: 7.14€.

	Predicted no CRC	Predicted CRC
No CRC	339 472.5 ± 4.0	16.54 ± 4.0
CRC	138.9 ± 4.6	79.1 ± 4.6

Table 17
Features of two new screening devices, Dev1 and Dev2.

Device	Cost	Specif.	Sensit.	Comfort
Dev1	250	0.85	0.8	2
Dev2	3	0.85	0.94	3

screening devices.[2] We discuss here how the approach proposed may be used to benchmark new devices. For illustration purposes, suppose we have come out with two new devices with features as in Table 17.

First of all, we should check whether the new devices are not dominated by the current ones. As an example, we would reject Dev1 because it is dominated by sDNA, whose features are $cost = 236.88€$, $specif = 0.866$, $sensit = 0.923$ and $comfort = 3$. However, Dev2 is non-dominated. Let us assess it with our model. Fig. 5 shows that for most of the range of the CRC probability below 0.55, the new test reduces uncertainty more than currently available screening tests. For extremely low CRC probability values, FIT still provides more information due to its better specificity. However, as Dev2 is much cheaper than FIT and has the same comfort value, we have that for the benchmark patient, the EU of the new test is 0.179, which is higher than that of the previous recommendation which was *no screening* (Section 3.1).

Suppose that we implement this test. As Table 18 shows, with the considerations that our model makes in terms of cost, comfort, and information, we would be recommending this test to the entire population. However, this is in general hard to achieve and would entail a large cost (63.86€ per patient) due to the high number of colonoscopies carried out. Hence, we generate a case in which we have 50,000 new Dev2 tests and the same resources as in the examples from Section 3.3. Table 19 provides the classification results with this strategy, which essentially increases the number of detected patients by more than 10, an increase of around 20%. The number of false positives also increases, as well as the total cost (9.85€ per patient), as more patients are being predicted CRC positive and require a colonoscopy. However, the increment in the F1 score [31], a common classification measure that balances the importance of true and false positives and negatives, from 0.50 in the original strategy to 0.54 shows that the benefit would be significant in classification terms.

3.5. Sensitivity analysis

It is important to assess the impact of the inputs on the output of the analysis through a sensitivity analysis. We focus here on the impact of the utility function obtained from the screening resource allocation. We restrict ourselves to the case of no resource limitation, as it may potentially reveal more variability.

3.5.1. Sensitivity analysis with respect to the probability equivalent

We first analyze the impact of the probability equivalent (PE) elicited to determine the risk aversion coefficient in the allocation of screening tests. Fig. 6 plots the optimal screening resource allocations showing how information generally impacts test allocation more than cost, as changes are more noticeable vertically than horizontally. In general, the total number of recommended tests increases as the weight given to the value of information increases and decreases as the weight given to cost increases.

The intuition behind this is that sDNA is, given the comfort levels, the recommended test for higher-risk patients as it is the most sensitive one and thus provides the highest quality information. However,

[2] As an example, this is one of the aims of the ONCOSCREEN project https://oncoscreen.health/.

Table 18
Comparison of strategies when adding the new test.

	Nothing	Colon.	gFOBT	FIT	Blood	sDNA	CTC	CC	New	Total
No lim. recom.	0	0	0	0	0	0	0	0	339 707	339 707
Op. limit	∞	3000	30 000	42 000	7000	5000	2000	2000	50 000	–
Final recom.	287 920	0	0	1787	0	0	0	0	50 000	51 787

Table 19
New strategy including new tests. Cost per patient: 9.85€.

	Predicted no CRC	Predicted CRC
No CRC	$339\,458.5 \pm 5.4$	30.5 ± 5.4
CRC	126.8 ± 3.8	91.2 ± 3.8

suppose a larger weight is given to information at lower costs. In that case, the number of FITs will increase as it is one of the cheapest methods and provides quite competitive information regarding CRC presence. Suppose now that information is valued more at higher costs. In that case, the number of sDNA tests allocated will increase as long as its information-cost tradeoff surpasses the utility of not undergoing screening.

3.5.2. Sensitivity analysis with respect to the comfort parameters

Let us discuss now sensitivity with respect to the comfort parameters λ. Remember that by repeating the elicitation exercise for comfort 3 screening methods, we ensured some robustness of the method in relation to λ_3. However, for the other comfort levels, this cannot be done due to the absence of additional tests for comparison. Moreover, the elicitation of comfort values depends on the available tests and does not account for alternative or future tests and their respective information values. This is complicates extrapolating results to new strategies, as seen in Section 3.4 where, with the established elicited parameters using the original tests, the new test is recommended to the entire population due to its great features, being this recommendation far from manageable. We look into how the variability in these values can lead to differences in recommendations and thus highlight the importance of consistency and robustness in the elicitation protocol.

Fig. 7 presents two screening recommendation allocations. The first one corresponds to decreasing the value of level 3 comfort to $\lambda_3 = 6.3$; observe how a lower comfort parameter reduces the number of screenings performed, increasing the amount of non-screened people. The second case corresponds to raising the value of level-1 comfort to $\lambda_1 = 4.8$ and level 2 to $\lambda_2 = 5$; the increase in the weight given to the comfort value for colonoscopy has a large effect in increasing the number of recommended sDNA tests. Recall that a colonoscopy is usually performed after a positive screening result, and thus, as the weight given to the comfort value for colonoscopy increases, the expected value of all screening methods also increases. However, this change is more noticeable for sDNA, it being the most sensitive tool, thus detecting more CRC positive cases.

Thus, a moderate modification of comfort parameters seems to impact program results. To ensure consistency, we would recommend following the elicitation protocol in a moderate number of patients chosen at random, and asking them about their indifference between cost, information, and comfort. Repetition and randomization would enable a more robust estimation of comfort parameters which would benefit model applicability.

4. Discussion

We have developed a decision analysis model and its accompanying decision support system for personalized screening in connection with CRC detection. Stemming from an earlier predictive model [15] we incorporated decision, value, and additional chance nodes, as well as new arcs, assessing the new probability tables. We introduced a multi-attribute utility model to support what combination of screening and colonoscopy decisions should be implemented for an individual with certain features. This could help strengthening the monitoring systems of colorectal screening programs, as has been proposed for European countries [32]. As showcased, this facilitates a more personalized approach to CRC screening within the general population as well as designing large-scale screening strategies taking into account constraints on treatment availability. Apart from the viability of the use of the model in public health management, it can also serve to simulate screening scenarios, regarding how to distribute screening tests according to different risk levels and estimate the costs of different strategies. The system enables as well benchmarking of novel screening devices.

The preference model incorporated was that in Section 2.2.2, and assimilated to that of the policy-makers designing or assessing a screening strategy. Its construction is based on preference elicitation techniques, a standard tool in decision analysis for parameter estimation. This may be easily modified, extending it to the case of several PMs and screening methods, and eventually can be individualized for patients to promote a more personalized approach to screening as suggested in the [33] guidelines.

As Section 3.1 illustrates, the model may be adapted and enhanced when additional variables are taken into account. In particular, this may be the case when data concerning diet, genetics or CRC family history are available, which was not the case in our initial dataset. The availability of information on additional variables can be very useful in improving the possibilities of better personalizing screening programs, giving place to more complex workflows for health systems which could include other types of medical-related models [34]. In any case, the major strength of this work lies in its development of a practical framework and methodology adaptable to more complex scenarios. As an example, we presented the assessment of the French screening strategy, but the framework may be replicated in other countries, especially when no screening programs are in place. Further, the model could be extended to consider a temporal setting, which could aid in modeling patient and disease behavior through time, and could leverage tools from common health technology assessment methods, such as, for example, measuring QALYs after performing screening or colonoscopy.

At least, two future lines emerge at least from this work. The first one refers to benchmarking the new screening devices that are being developed within the ONCOSCREEN project with the methodology specified in Section 3.4. The second one, following the above mentioned [3], which found statistical evidence that less participation in screening programs corresponds to unmarried adult patients with a low educational level, who live in rural areas, have an unhealthy lifestyle or spend long periods without health checks, stems from the low acceptance of CRC screening programs in many countries, and will focus on extending our model with incentive mechanisms to promote screening adoption. These will involve modeling variables regarding acceptability and health literacy, which highly influence screening uptake.

CRediT authorship contribution statement

Daniel Corrales: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **David Ríos Insua:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Formal analysis. **Marino J. González:** Writing – review & editing, Validation, Supervision, Project administration.

Ethics statement

All authors agree that:

This research presents an accurate account of the work performed, all data presented are accurate and methodologies detailed enough to permit others to replicate the work.

This manuscript represents entirely original works and or if work and/or words of others have been used, that this has been appropriately cited or quoted and permission has been obtained where necessary.

This material has not been published in whole or in part elsewhere.

The manuscript is not currently being considered for publication in another journal.

All authors have been personally and actively involved in substantive work leading to the manuscript and will hold themselves jointly and individually responsible for its content.

Declaration of Generative AI and AI-assisted technologies in the writing process

That generative AI and AI-assisted technologies have not been utilized in the writing process or if used, disclosed in the manuscript the use of AI and AI-assisted technologies and a statement will appear in the published work.

That generative AI and AI-assisted technologies have not been used to create or alter images unless specifically used as part of the research design where such use must be described in a reproducible manner in the methods section.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the AXA-ICMAT Chair in Adversarial Risk Analysis; the Spanish Ministry of Science project PID2021-124662OB-I00; and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement N. 101097036 (ONCOSCREEN). D. Corrales is hired under the Generation D initiative, promoted by Red.es, an organisation attached to the Ministry for Digital Transformation and the Civil Service, for the attraction and retention of talent through grants and training contracts, financed by the Recovery, Transformation and Resilience Plan through the European Union's Next Generation funds. The authors thank Quirónprevención for the provision of data.

Appendix A. Variables in the original BN model

See Table 20.

Appendix B. Probabilities available from public sources

To assess the probabilities in the chance nodes introduced, Table 21 compiles the performance information of the screening methods as well as their cost with data available from Barré et al. [5], based on French sources.

We also require the probabilities and expected costs of complications as reflected in Table 22 based on [5].

Appendix C. Full elicitation of comfort parameters

Table 23 contains full details of the calculations used to assess comfort parameters.

Table 20
Fourteen variables in the BN model.

Variable	Definition	Levels
v_{sex}	Sex	{*female, male*}
v_{age}	Age	(24,34], (34,44], (44,54], (54,64]
v_{SES}	Socioeconomic status	{1,2,3}
v_{BMI}	Body mass index	{*underw., normal, overw., obese*}
v_{PA}	Physical activity	{*insufficiently active (1), sufficiently active (2)*}
v_{SD}	Sleep duration	{*short, normal, excessive*}
v_{alc}	Alcohol consumption	{*low, high*}
v_{smok}	Smoker profile	{*non-smoker, ex-smoker, smoker*}
v_{anx}	Anxiety	{*yes, no*}
v_{dep}	Depression	{*yes, no*}
v_{hypten}	Hypertension	{*yes, no*}
$v_{hypchol}$	Hypercholesterolemia	{*yes, no*}
v_{diab}	Diabetes	{*yes, no*}
v_{CRC}	Colorectal cancer	{*yes, no*}

Table 21
Specificity, sensitivity, and cost in€ of interventions.

	gFOBT	FIT	BldBsd	sDNA	CTC	CC	Colons.
Sensitivity	0.45	0.75	0.66	0.923	0.8	0.87	0.97
Specificity	0.978	0.966	0.91	0.866	0.89	0.92	0.99
Cost€	12.14	14.34	123.13	236.88	95.41	510.24	1000

Appendix D. Information assessment

This appendix provides full details of the information assessment procedure employed to develop our value function.

Assume that we want to assess the amount of uncertainty reduced by a screening strategy. Let us denote by CRC the random variable describing the presence of CRC, and R_s and R_c the respective variables describing the results of screening and colonoscopy. Our interest is in calculating $MI(CRC;(R_s,R_c))$ where MI designates the *mutual information (MI)* function [35]. Both R_s and R_c depend on the decision of which screening test to use and whether or not to perform a colonoscopy, while the distribution of CRC will depend on the evidence collected from its parent nodes in the ID. However, we shall not make explicit such dependence to lighten the notation. Then, the mutual information is written as

$$\begin{aligned} MI(CRC;(R_s,R_c)) &= MI(CRC;R_s) + MI((CRC;R_c)|R_s) \\ &= \mathbb{E}_{crc,r_s}\left[\log\left(\frac{p(crc,r_s)}{p(crc)p(r_s)}\right)\right] \\ &\quad + \mathbb{E}_{crc,r_s,r_c}\left[\log\left(\frac{p(crc,r_c|r_s)}{p(crc|r_s)p(r_c|r_s)}\right)\right] \\ &= \mathbb{E}_{crc,r_s,r_c}\left[\log\left(\frac{p(crc|r_s)}{p(crc)}\right)\right. \\ &\quad \left.+\log\left(\frac{p(crc|r_s,r_c)}{p(crc|r_s)}\right)\right]. \end{aligned} \tag{3}$$

As the suggested decision in our problem will be based on maximum expected utility (hypothesis 4), a natural value function that will effectively quantify the information provided by a screening strategy is

$$pmi(crc,r_s,r_c) = \log\left(\frac{p(crc|r_s)}{p(crc)}\right) + \log\left(\frac{p(crc|r_s,r_c)}{p(crc|r_s)}\right),$$

which we designate the *pointwise mutual information* of CRC and (R_s,R_c). Although the domain of this function is the set of real numbers, the *MI* (that is, the expectation of the *PMI*) will always be positive and bounded by the entropy of either of the variables. As we are interested in the uncertainty concerning the presence of CRC, we normalize the *MI* dividing it by its entropy $H(CRC) = -\sum p(crc)\log p(crc)$, so that the final value function lies in the interval $[0,1]$. In summary, the value of information v_{info} provided by a screening strategy concerning the

Table 22
Probabilities and expected costs of complications for CRC interventions.

	gFOBT	FIT	BlbBsd	sDNA	CTC	CC	Colons	Cost
None	1	1	1	1	0.9996	0.9997	0.998	0€
Bleed.	0	0	0	0	0	0	0.0006	1241€
Reten.	0	0	0	0	0	0.0003	0	1241€
Perfor.	0	0	0	0	0.0004	0	0.001	2810€
Other	0	0	0	0	0	0	0.0004	6621€

Table 23
Eliciting parameter λ_k. X indicates the preferred alternative in the corresponding pairwise comparison.

Comfort	Scr. method	Cost	Info	Preference	Indiff. cost	$\hat{\lambda}_k$	λ_k
1	Colonos	1000	0.530		–	$\lambda_1 = 4.01$	$\lambda_1 = 4.01$
	Synth.	–	0.4	×	300€		
2	CTC	95.41	0.159	×	–	$\lambda_2 = 4.17$	$\lambda_2 = 4.17$
	CC	510.24	0.225		180€		
3	gFOBT	12.14	0.129		3€	$\lambda_3 = 5.04$	$\bar{\lambda}_3 = 6.80$
	FIT	14.34	0.245	×	–		
3	gFOBT	12.14	0.128	×	–	$\lambda_3 = 10.57$	
	Blood test	125.13	0.121		10€		
3	gFOBT	12.14	0.128	×	–	$\lambda_3 = 16.28$	
	sDNA	236.88	0.197		170€		
3	FIT	14.34	0.244	×	–	$\lambda_3 = 6.40$	
	Blood test	125.13	0.121		1.5€		
3	FIT	14.34	0.244	×	–	$\lambda_3 = 7.2$	
	sDNA	236.88	0.197		6€		
3	Blood test	125.13	0.121		80€	$\lambda_3 = 6.17$	
	sDNA	236.88	0.197	×	–		
4	No scr.	0	0	–	–	–	$\lambda_4 = 7$

presence of CRC will be given by the *normalized* or *relative pointwise mutual information (RPMI)* defined through

$$v_{info}(crc, r_s, r_c) = pmi(crc, r_s, r_c) \;/\; H(CRC). \tag{4}$$

Intuitively, the expected value of this function refers to the proportion of uncertainty reduced by the screening strategy from the total uncertainty in relation to the presence of CRC.

Tables 4 and 5 respectively provide the v_{info} for our benchmark patient when not undertaking screening and when performing FIT. Positive values correspond to correct predictions; negative values, to wrong ones; and, finally, zero indicates no change in uncertainty. Values corresponding to cases in which CRC is present are larger in magnitude as they have a low probability. This conforms with the argument that a missed CRC positive scenario is much worse than a misdiagnosed healthy patient [26].

Recall that we look for the best screening strategy and contemplate the decision to perform the colonoscopy depending on the output of the first screening. Thus, the value tables will not contain the expected v_{info} but rather the extension of its values depending on the results of screening, that is, $\mathbb{E}_{crc,r_c|r_s}[v_{info}]$, which does not necessarily lie in the $[0, 1]$ interval, as Table 6 portrays. Indeed, observe how for the first alternative regarding no screening, the value lies in $[0, 1]$ as no other result is possible and, thus, coincides with its expectation. When screening is performed, not much information is gained when the result predicts *false*, the most likely scenario in general. However, the information provided is much larger when the prediction is *true*, increasing its value when a colonoscopy is further performed. Notice how in a low prevalence scenario highly specific tests, like FIT, provide more information, whereas highly sensitive tests, like sDNA, do not have as much of an impact.

Fig. 3 plots the v_{info} function for all possible values of the probability of having CRC facilitating comparison of screening methods in terms of information. Observe that blood-test, CTC, and gFOBT are bounded above by the rest of screening methods in terms of the information measure used. From this, we conclude that we could discard blood-test and CTC as they are equal or worse to FIT in the four criteria *(information, comfort, specificity, sensitivity)* considered. This is not the case, however, for gFOBT as it is the cheapest method and has non-dominated performance metrics. Further observe that at very low probabilities of CRC, FIT is the method that provides the best information on its own.

Note that the asymmetries of v_{info} in Fig. 3 derive from the differences between specificity and sensitivity. More specific tests, like FIT, take their maximum at lower $p(CRC)$ values, whereas more sensitive ones, like sDNA, do so at larger $p(CRC)$ values.

References

[1] F. Kastrinos, S.S. Kupfer, S. Gupta, Colorectal cancer risk assessment and precision approaches to screening: Brave new world or worlds apart? Gastroenterology (ISSN: 0016-5085) 164 (5) (2023) 812–827, http://dx.doi.org/10.1053/j.gastro.2023.02.021, Personalizing GI Cancer Risk Assessment and Management: The Future is Now.

[2] A. Ferrari, I. Neefs, S. Hoeck, M. Peeters, G. Van Hal, Towards novel non-invasive colorectal cancer screening methods: a comprehensive review, Cancers 13 (8) (2021) 1820, http://dx.doi.org/10.3390/cancers13081820.

[3] I. Ola, R. Cardoso, M. Hoffmeister, H. Brenner, Utilization of colorectal cancer screening tests across European countries: a cross-sectional analysis of the European health interview survey 2018–2020, Lancet Reg. Health– Eur. 41 (2024) http://dx.doi.org/10.1016/j.lanepe.2024.100920.

[4] H. Sung, R.L. Siegel, M. Laversanne, C. Jiang, E. Morgan, M. Zahwe, Y. Cao, F. Bray, A. Jemal, Colorectal cancer incidence trends in younger versus older adults: an analysis of population-based cancer registry data, Lancet Oncol. (2024) http://dx.doi.org/10.1016/S1470-2045(24)00600-4.

[5] S. Barré, H. Leleu, R. Benamouzig, J.-C. Saurin, A. Vimont, S. Taleb, F. De Bels, Cost-effectiveness analysis of alternative colon cancer screening strategies in the context of the French national screening program, Ther. Adv. Gastroenterol. 13 (2020) 1756284820953364, http://dx.doi.org/10.1177/1756284820953364.

[6] European Colorectal Cancer Screening Guidelines Working Group, et al., European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication, Endoscopy 45 (01) (2013) 51–59, http://dx.doi.org/10.1055/s-0032-1325997.

[7] E. Ferlizza, R. Solmi, M. Sgarzi, L. Ricciardiello, M. Lauriola, The roadmap of colorectal cancer screening, Cancers 13 (5) (2021) 1101, http://dx.doi.org/10.3390/cancers13051101.

[8] A. Fouladi, A. Asadi, E.A. Sherer, M. Madadi, Cost-effectiveness analysis of colorectal cancer screening strategies using active learning and Monte Carlo simulation, Med. Decis. Making 44 (5) (2024) 554–571, http://dx.doi.org/10.1177/0272989X241258224.

[9] G.B. Taksler, A.T. Perzynski, M.W. Kattan, Modeling individual patient preferences for colorectal cancer screening based on their tolerance for complications risk, Med. Decis. Making 37 (3) (2017) 204–215, http://dx.doi.org/10.1177/0272989X16679161.

[10] F. Jensen, T. Nielsen, Bayesian Networks and Decision Graphs, second ed., Springer, New York, ISBN: 9780387915029, 2007, http://dx.doi.org/10.1007/978-0-387-68282-2.

[11] R.D. Shachter, Evaluating influence diagrams, Oper. Res. 34 (6) (1986) 871–882, http://dx.doi.org/10.1287/opre.34.6.871.

[12] L. Souza-Pereira, N. Pombo, S. Ouhbi, V. Felizardo, N. Garcia, Clinical decision support systems for chronic diseases: A systematic literature review, Comput. Methods Programs Biomed. 195 (2020) 105565.

[13] J.G. Dolan, E. Boohaker, J. Allison, T.F. Imperiale, Can streamlined multicriteria decision analysis be used to implement shared decision making for colorectal cancer screening? Med. Decis. Making 34 (6) (2014) 746–755, http://dx.doi.org/10.1177/0272989X13513338.

[14] E.H. van den Bogaart, M.E. Kroese, M.D. Spreeuwenberg, D. Ruwaard, A. Tsiachristas, Economic evaluation of new models of care: does the decision change between cost-utility analysis and multi-criteria decision analysis? Value Heal. 24 (6) (2021) 795–803.

[15] D. Corrales, A. Santos-Lozano, S. López-Ortiz, A. Lucia, D.R. Insua, Colorectal cancer risk mapping through Bayesian networks, Comput. Methods Programs Biomed. (2024) 108407, http://dx.doi.org/10.1016/j.cmpb.2024.108407.

[16] BayesFusion, LLC, GeNIe modeler: Complete modeling freedom, 2024, https://support.bayesfusion.com/docs/GeNIe.pdf, (Accessed 10 February 2025).

[17] D. Corrales, Decision_model_screening_crc, 2024, URL https://github.com/DanielCorralesAlonso/Decision_Model_Screening_CRC.

[18] S. French, D. Rios Insua, Statistical Decision Theory, Wiley, 2000.

[19] EUnetHTA Joint Action 2, Work Package 7, Subgroup 3, E. Heintz, A. Gerber-Grote, S. Ghabri, F.F. Hamers, V.P. Rupel, R. Slabe-Erker, T. Davidson, Is there a European view on health economic evaluations? Results from a synopsis of methodological guidelines used in the EUnetHTA partner countries, Pharmacoeconomics 34 (2016) 59–76, http://dx.doi.org/10.1007/s40273-015-0328-1.

[20] R.L. Keeney, R.S. Gregory, Selecting attributes to measure the achievement of objectives, Oper. Res. 53 (1) (2005) 1–11, http://dx.doi.org/10.1287/opre.1040.0158.

[21] G.W. Torrance, D.H. Feeny, W.J. Furlong, R.D. Barr, Y. Zhang, Q. Wang, Multiattribute utility function for a comprehensive health status classification system: Health utilities index mark 2, Med. Care 34 (7) (1996) 702–722, http://dx.doi.org/10.1097/00005650-199607000-00004.

[22] M.B. Westover, N.A. Eiseman, S.S. Cash, M.T. Bianchi, Information theoretic quantification of diagnostic uncertainty, Open Med. Inform. J. 6 (2012) 36.

[23] W.A. Benish, A review of the application of information theory to clinical diagnostic testing, Entropy 22 (1) (2020) 97, http://dx.doi.org/10.3390/e22010097.

[24] A. Casagrande, F. Fabris, R. Girometti, Fifty years of Shannon information theory in assessing the accuracy and agreement of diagnostic tests, Med. Biol. Eng. Comput. 60 (4) (2022) 941–955.

[25] R.R. Love, A.E. Camilli, The value of screening, Cancer 48 (S1) (1981) 489–494, http://dx.doi.org/10.1002/1097-0142(19810715)48:1+<489::AID-CNCR2820481309>3.0.CO;2-F.

[26] J.P. Ioannidis, R. Tarone, J.K. McLaughlin, The false-positive to false-negative ratio in epidemiologic studies, Epidemiology 22 (4) (2011) 450–456, http://dx.doi.org/10.1097/EDE.0b013e31821b506e.

[27] J. González-Ortega, V. Radovic, D. Ríos Insua, Utility elicitation, in: Elicitation: The Science and Art of Structuring Judgement, Springer, 2018, pp. 241–264, http://dx.doi.org/10.1007/978-3-319-65052-4_10.

[28] S. French, Decision Theory: An Introduction to the Mathematics of Rationality, Halsted Press, 1986, doi:abs/10.5555/27012.

[29] P.H. Farquhar, State of the art—utility assessment methods, Manag. Sci. 30 (11) (1984) 1283–1300, http://dx.doi.org/10.1287/mnsc.30.11.1283.

[30] A. Pellat, J. Deyra, R. Coriat, S. Chaussade, Results of the national organised colorectal cancer screening program with FIT in Paris, Sci. Rep. 8 (1) (2018) 4162, http://dx.doi.org/10.1038/s41598-018-22481-9.

[31] C.J. Van Rijsbergen, Information Retrieval. 2nd. Newton, Ma, USA: Butterworth-Heinemann, 1979.

[32] E. Altobelli, L. Rapacchietta, C. Marziliano, G. Campagna, V.F. Profeta, R. Fagnano, Differences in colorectal cancer surveillance epidemiology and screening in the who European region, Oncol. Lett. 17 (2) (2019) 2531–2542, http://dx.doi.org/10.3892/ol.2018.9851.

[33] R.J. Volk, V.B. Leal, L.E. Jacobs, A.M. Wolf, D.D. Brooks, R.C. Wender, R.A. Smith, From guideline to practice: new shared decision-making tools for colorectal cancer screening from the American cancer society, CA: Cancer J. Clin. 68 (4) (2018) 246–249, http://dx.doi.org/10.3322/caac.21459.

[34] Y. Wang, J. Zhao, H. Xu, C. Han, Z. Tao, D. Zhou, T. Geng, D. Liu, Z. Ji, A systematic evaluation of computational methods for cell segmentation, Brief. Bioinform. 25 (5) (2024) bbae407.

[35] T.M. Cover, Elements of Information Theory, John Wiley & Sons, 1999, http://dx.doi.org/10.1002/047174882X.