A decision analysis model for colorectal cancer screening

Daniel Corrales, MSc, *daniel.corrales @ icmat.es* [*a], David Ríos Insua, PhD, *david.rios @ icmat.es* [a], and Marino J. González, PhD, *marino-jose.gonzalez @ unirioja.es* [b]

[a]*Institute of Mathematical Sciences, ICMAT-CSIC, 28049 Madrid, Spain*
[b]*Department of Economics, University of La Rioja, 26006 Logroño, Spain*

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, supporting decisions concerning whether and which screening method to consider and/or whether a colonoscopy should be administered.

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

Results   The proposed model is used to support personalized individual screening based on relevant characteristics. It serves to assess existing national screening programs and design new ones. In particular, it suggests replacing current age-based strategies followed in many European countries by more personalized strategies based on the type of model proposed. Additionally, the model facilitates benchmarking of novel screening devices.

Conclusions   This work creates a framework supporting personalized CRC screening improving upon current age-based screening strategies.

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] focused on age and using fecal testing and colonoscopy (CS). The latter is considered highly invasive, influencing quite negatively program uptake among the general population [2]. It has been reported [3] through the European Health

---

*Corresponding author.

Interview Survey (2018-2020) that the coverage of faecal tests in the population aged 50-74 varies according to the organisation of the screening programme. In countries with fully rolled-out programmes it varies from 37.7% in Croatia to 74.9% in Denmark; in turn, in countries without programmes or with localised programmes it varies from 6.3% in Bulgaria to 34.2% in Latvia. This study also found statistical evidence that less participation in screening programmes corresponds to unmarried adult patients which have a low educational level, live in rural areas, have an unhealthy lifestyle or spend long periods without health checks. 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 the disease and 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 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 decisions under uncertainty. These analyses have proved to be relevant in medical decision-making [12, 13].

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 [14] and complement it to design an ID that identifies the screening methods adopted, their impacts, and their associated utilities. It will be used 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, assessing national screening strategies, designing such strategies, and specifying benchmarks for novel screening technologies. Software supporting the proposed approach produced in GeNIe [15] and its Python wrapper PySMILE is available at [16] 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 relies on a BN predictive model previously developed for CRC risk mapping purposes [14] . The BN aggregates exhaustive expert information and data from a large occupational health assessment study. It relates modifiable (physical activity, sleep duration, alcohol consumption, smoking status, body mass index (BMI), anxiety, depression) and non-modifiable (sex, age, socio-economic status) risk factors as well as medical conditions (hypercholesterolemia, hypertension, diabetes) relevant to CRC, assessing such relations through local probability distributions at the nodes. Figure 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 a large database of annual occupational health assessments (2M people). Procedures to assess the probability tables and the BN use for risk mapping and influential variable detection are detailed in [14].

2.1.2 CRC screening decision model

For the construction of the ID for the CRC screening decision-support problem, a set of variables (chance, decision, and values) are defined and incorporated, together with the corresponding arcs, into the original BN.

Hypotheses

Before presenting the variables, let us first introduce four hypotheses 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.
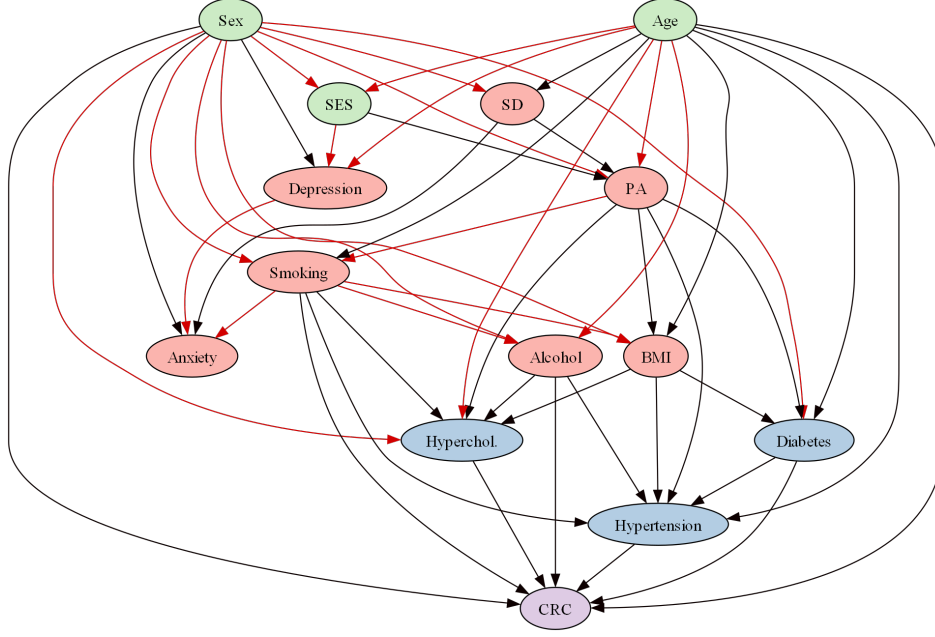
Figure 1: Originating CRC Bayesian Network [14]

2. We adapt to the standard strategy 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 to be 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. Note that, in doing this, we focus on the short-term outcomes of the screening intervention, ignoring pointers to longer-term outcomes like, e.g., expected QALYs gained. This is motivated mostly by our interest in benchmarking novel screening devices 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 decisions will be based on the maximum expected utility (EU) principle [17], with the utility vision of either the patient, the doctor, or the health policy maker, as later discussed.

These hypotheses are broadly aligned with the recommendations from the European Network for Health Technology Assessment [18], reflecting a shared direction while allowing for nuanced differences. Based on them, we describe the remaining elements required to structure qualitatively the problem.

New nodes and arcs

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

4

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 test*, *computed tomography colonography* (CTC) and *colon capsule* (CC). Of course, 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 information, beyond age, is used to support the corresponding screening decisions.

Chance nodes. Besides the chance nodes from the original BN, we include three additional ones. First, we consider the potential *complications* associated with the eventual screening and colonoscopy interventions [5], which are *bleeding*, *retention*, *perforation*, *death* and, obviously, *no complications.* The other two nodes refer to the results of both interventions, *screening results* and *colonoscopy results*, 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 [17] to support CRC screening decisions. As primary criteria, the model includes the *cost of complications*, the *cost of the intervention*, the *comfort of the intervention* 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.

Arcs 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 of all, based on hypothesis 1, initially we 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), hence the arcs from such variables to 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 facilitate covering all standard screening protocols. 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 coming from results of screening, results of colonoscopy, and CRC.

2.1.3 Final influence diagram

Figure 2 reflects the structure of the ID produced. In line with our comments, the possible 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. The diagram was presented to several medical experts who validated it for concept and meaning.
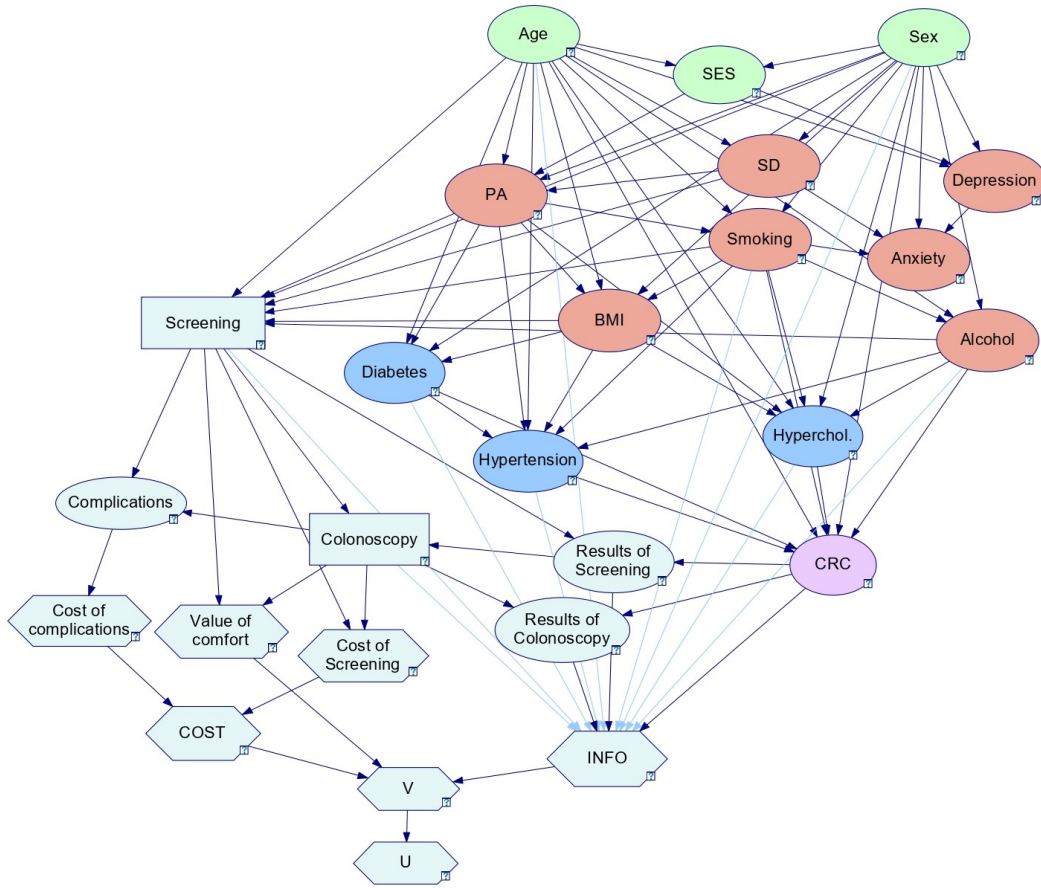

Figure 2: CRC Screening Decisions Influence Diagram

2.2 Quantifying the CRC screening decision model

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

2.2.1 Probability models at chance nodes

First, the ID inherits the probability models at nodes from the original BN. We refer to [14] for details concerning how the corresponding tables were built based on prior distributions and the available occupational health assessment database. Note that, as a consequence, the probabilities available were assessed with some uncertainty through posterior predictive distributions. We shall 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 19 (Appendix A) includes the sensitivities and specificities of screening methods and that of colonoscopy, from which we deduce the required probabilities in the nodes *results of screening* and *results of colonoscopy*, 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 presence of CRC and the use of colonoscopy would be Table 1, with e.g. the probability of reporting *Doesn't have CRC*, given that the person goes through a colonoscopy and she does not have CRC, being 0.03.

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

Table 1: Probabilities of colonoscopy results, given antecessors.

Similarly, the probabilities in the first seven columns of Table 20 (Appendix A) 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 [16].

2.2.2 Preference models

We discuss first the preference model for each criterion and 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 € assessed in [5] for France, displayed in Table 19 (bottom row).

Cost c_{comp} of complications. These are assessed through their entailed expected costs in € and obtained from public sources in [5] for France, available in Table 20 (right column).

Comfort *com*. Lacking a natural attribute to assess intervention comfort, we used a *constructed attribute* [19] 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.

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 slight 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 2: Comfort levels for interventions.

Torrance et al. [20] 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 the quality of life over the years by creating a utility function, whereas ours, in line with a short-term health assessment focus [18], is more geared towards just 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 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. Further, the patient 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 FIT with positive result would move the individual from 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).

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

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

Therefore, we seek ways to evaluate such changes in uncertainty, that is to assess the value of the information that the interventions provide concerning the presence of CRC.

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 [21]. Note that 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

$$MI(CRC; (R_s, R_c)) = MI(CRC; R_s) + MI((CRC; R_c)|R_s) \tag{1}$$
$$= \mathbb{E}_{crc,r_s}\left[\log\left(\frac{p(crc, r_s)}{p(crc)p(r_s)}\right)\right] + \mathbb{E}_{crc,r_s,r_c}\left[\log\left(\frac{p(crc, r_c|r_s)}{p(crc|r_s)p(r_c|r_s)}\right)\right]$$
$$= \mathbb{E}_{crc,r_s,r_c}\left[\log\left(\frac{p(crc|r_s)}{p(crc)}\right) + \log\left(\frac{p(crc|r_s, r_c)}{p(crc|r_s)}\right)\right].$$

Now, 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 would be

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

We refer to this value as 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 shall 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 presence of CRC will be given by the *normalized* or *relative pointwise mutual information (RPMI)* defined through

$$v_{info}(crc, r_s, r_c) = pmi(crc, r_s, r_c) \,/\, H(CRC). \tag{2}$$

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.

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 [22] 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

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

Table 4: No screening 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 5: FIT v_{info} for benchmark.

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 [23]. 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.

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 6: Expected information of different policies for reference patient

Figure 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 Figure 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.
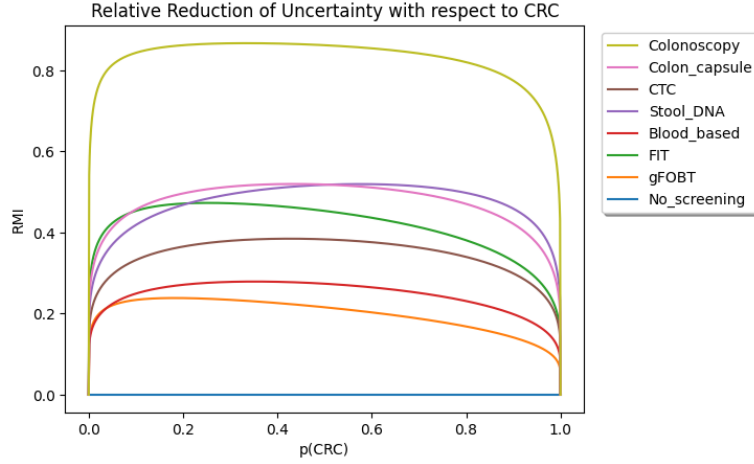
Figure 3: v_{info} for various CRC screening alternatives.

Multiple criteria preference aggregation

Lastly, let us aggregate the criteria through a utility function u, first employing a multicriteria value function, then transformed to take into account risk aversion, see [24] 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 [24], 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, adopting the model

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

where λ_k 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 λ_k parameters is delicate and 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 both cheaper and more informative than the other, and that the individual declaring his preferences reveals that he favours $(info_1, cost_1)$ to $(info_2, cost_2)$, represented as $(info_1, cost_1) \succ (info_2, cost_2)$. Then, we interactively ask the individual for a $\overline{cost}$ value smaller than $cost_2$ such that $(info_1, cost_1) \sim (info_2, \overline{cost})$. Such options would have the same value, that is,

$$\lambda_k \times info_1 - \log_{10}(cost_1 + 1) = \lambda_k \times info_2 - \log_{10}(\overline{cost} + 1), \tag{3}$$

and we would 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. [25]. 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 and robustify 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. Note that we would expect that for lower discomfort, the value of λ should be larger, that is, $\lambda_4 > \lambda_3 > \lambda_2 > \lambda_1$, suggesting that information is more valuable when obtained from a more comfortable screening tool.

Example 1 *For reference purposes, consider the benchmark patient to elicit the comfort parameters. Table 7 presents the v_{info} provided by all interventions. Figure 4*

	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

Table 7: Expected information of interventions benchmark

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 for examples.*
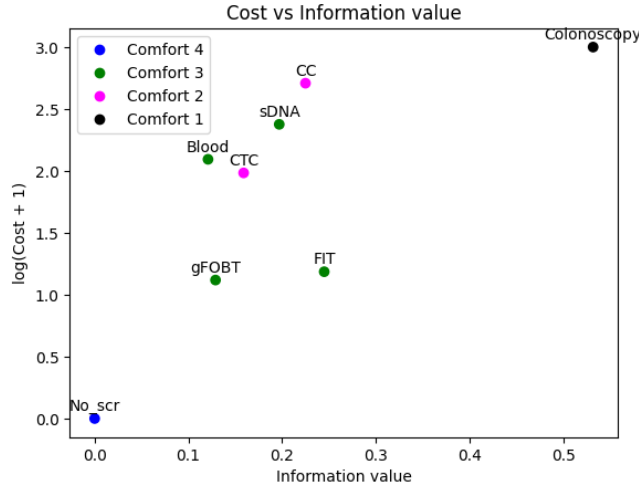

Figure 4: Scatterplot

To illustrate the elicitation process, let us focus on comfort level 2. CTC and CC respectively provide a v_{info} of 0.159 and 0.225, while their aggregated costs are 96.51€ and 510.64€. Suppose the individual declares preferring CTC to CC. He is interactively asked how much would the cost of CC have to be so that they are indifferent to one or the other. Assuming that the interview converges to 180€, using (3) 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 21 (Appendix B) reflects the full elicitation exercise.

Parameter	Value
λ_1	4.01
λ_2	4.17
λ_3	6.80
λ_4	7

Table 8: Values of λ parameters.

Comfort levels 1 and 4 are slightly different as we only have one option for them. For comfort level 1 we 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, in this case both the v_{info} and the 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.

$\triangle$

Once the value function is elicited, assuming (constant absolute) risk aversion [24] we adopt the following expression for the utility function

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

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 [26].

Example 2 *(Cont.) Observe that, 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, the worst outcome in terms of information would be not doing anything as it has no added value. Concerning cost, the best possible option would be a costless test, while the worst possible option would be a very expensive test with all possible complications, having an added cost for the patient. Thus, we choose as references the pairs ($cost^* = 0\text{€}$, $info^* = 15.75$) and ($cost_* = 8131.71\text{€}$, $info_* = 0$).*

Suppose now that the reference to assess the risk aversion coefficient is ($cost = 50\text{€}$, $info = 4.1$).[1] Assume all interventions are comfort level 3 and suppose we interactively obtain from the interview that the PE is 0.7. We then have the system

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

leading to the parameters in Table 9.

$\triangle$

We shall employ such parameters in our discussion of use cases in Section 3.

---

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

Param.	Value
a	1.015
b	0.872
ρ	0.039

Table 9: Utility parameters.

3 Results

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

3.1 Managing individual cases

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 summarised in Table 10.

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-based, are preferred to no screening. Intuitively, in the context of 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 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 cost of a colonoscopy. Indeed, the information provided by 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.

	1st recom	EU	2nd recom	EU	EU Pred False No col	Col	EU Pred True 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

Table 10: Personalized screening strategies in the four cases

3.2 Assessing the French national screening strategy

Our goal now is to assess, as an example, the current *one-size-fits-all* French screening strategy based on age. 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 [27]. Let us present three cases that are not properly prioritized in the current strategy.

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 required 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 maximum EU method is FIT with 0.143, followed by a colonoscopy if its result is positive.

These examples show how the current age-based strategy may fail in various ways at detecting cases with highest EU. 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 aid in redistributing the efforts for more efficient screening programs, as we next show.

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 350000 individuals whose records were kept for testing purposes for the BN model in Section 2.1.1, as fully described in [14]. We use all thirteen available variables in the network in this use case. As our

database does not include data from screening or colonoscopy results, we simulate them based on the sensitivity and specificity information of various interventions available in Table 19.

When designing a screening program, we should expect constraints on 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 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 intervention choice will be that providing highest EU to the individual; once we saturate the n_i tests available for the i-th screening method, we remove it from the list of available methods and search for those of maximum EU among the remaining ones. This 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.This is a reasonable assumption as it is the standard health practice.

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

1. We assume there are no constraints when applying this 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 direct colonoscopies to 3000; gFOBT to 30000; FIT to 42000; blood-based tests to 7000; sDNA to 6000; CTC to 2000; and, CC to 2000, therefore being able to cover 91000 individuals (out of the 350000).

3. Finally, the current national age-based strategy.

With the parameters developed throughout the text, the distribution of recommended tests and results for the first two strategies would be as Table 11 shows, where the first row indicates tests administered when no constraints are included; the middle row indicates test constraints; and, finally, the last row indicates the number of tests when constraints are included. Interestingly, only sDNA and FIT are the recommended strategies. 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 291707 patients, no screening is preferred to the remaining alternative screening methods even when setting operational limits.

	Nothing	Colon.	gFOBT	FIT	Blood	sDNA	CTC	CC	Total
No lim recom.	291323	0	0	560	0	47824	0	0	48384
Op. limit	∞	3000	30000	42000	7000	6000	2000	2000	-
Final recom.	291707	0	0	42000	0	6000	0	0	48000
National	290633	0	0	49074	0	0	0	0	49074

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

We apply the three strategies described and obtain the results for 200 simulations. Tables 12-14 contain their confusion matrices for CRC classification with values corresponding to the mean over the 200 simulations and their standard deviations. Table

12 shows how the current age-based strategy has a poor performance in detecting patients with CRC, that is a poor sensitivity (0.36), while the proposed unconstrained new strategy, shown in 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 for the proposed strategy. Concerning cost, the unconstrained new strategy entails an increase in the average cost (44.62 € vs. 7.24€ per patient). To account for that excessive increase in cost and the considerable precision decrease, we set operational limits (middle row 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 the sensitivity (0.37) and a decrease in precision (0.79), but both are more subtle now. 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.

	Predicted No CRC	Predicted CRC
No CRC	339472.1 ± 4.0	16.9 ± 4.0
CRC	139.8 ± 4.4	78.3 ± 4.4

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

	Predicted No CRC	Predicted CRC
No CRC	339425.2 ± 8.2	63.8 ± 8.2
CRC	121.3 ± 3.1	96.7 ± 3.1

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

	Predicted No CRC	Predicted CRC
No CRC	339466.5 ± 4.9	22.5 ± 4.9
CRC	134.8 ± 4.1	83.2 ± 4.1

Table 14: Mean classification results for new strategy with operational constraints. Cost per patient: 12.81€

For validation purposes, assume that the current system can cover 49,074 FIT tests, the number of people in the population from our 2016 dataset in the age range [54,64]. Assuming this, we compare the current strategy with our model by taking the 49,074 patients with highest FIT EU and performing 200 simulations to assess the differences in performance and cost. Table 15 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 33000€. Although the improvement might seem relatively low, when extrapolated to a real-sized population, the method can save many lives and money. As an example, France and Spain respectively have

populations of $8,591,286$ and $6,583,183$ within the screening age [54-64]. Thus, just an implementation of the model within the mentioned target population would detect around 175 and 134 more positive patients in France and Spain, respectively.

	Predicted No CRC	Predicted CRC
No CRC	339472.5 ± 4.0	16.54 ± 4.0
CRC	138.9 ± 4.6	79.1 ± 4.6

Table 15: Mean classification results for new strategy on 49074 patients with highest FIT utility. Cost per patient: 7.14€

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 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 16.

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

Table 16: Features of two new screening devices, Dev1 and Dev2

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. Figure 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).

	Nothing	Colon.	gFOBT	FIT	Blood	sDNA	CTC	CC	New	Total
No lim. recom.	0	0	0	0	0	0	0	0	339707	339707
Op. limit	∞	3000	30000	42000	7000	5000	2000	2000	50000	-
Final recom.	287920	0	0	1787	0	0	0	0	50000	51787

Table 17: Comparison of strategies when adding the new test.

Suppose that we implement this test. As Table 17 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
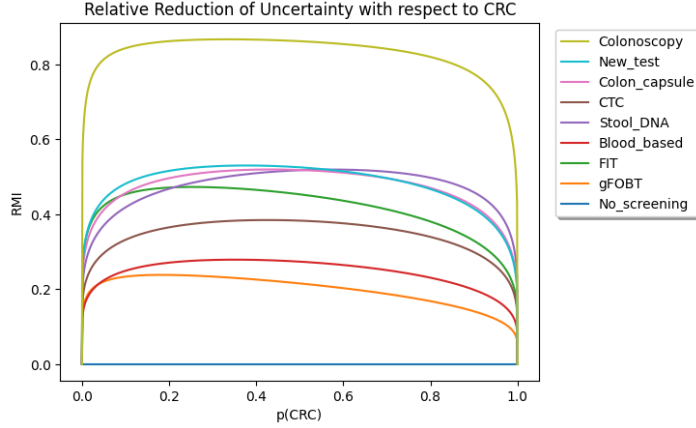

Figure 5: Plot of v_{info} at different levels of $p(CRC)$

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 18 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, a common classification measure that balances the importance of true and false positives and negatives [28], from 0.50 in the original strategy to 0.54 shows that the benefit would be significant in classification.

	Predicted No CRC	Predicted CRC
No CRC	339458.5 ± 5.4	30.5 ± 5.4
CRC	126.8 ± 3.8	91.2 ± 3.8

Table 18: New strategy including new tests. Cost per patient: 9.85€

3.5 Sensitivity analysis

We conclude by performing several relevant sensitivity analysis with our model, specifically in relation to the utility function obtained.

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. Figure 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, 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 performing screening.
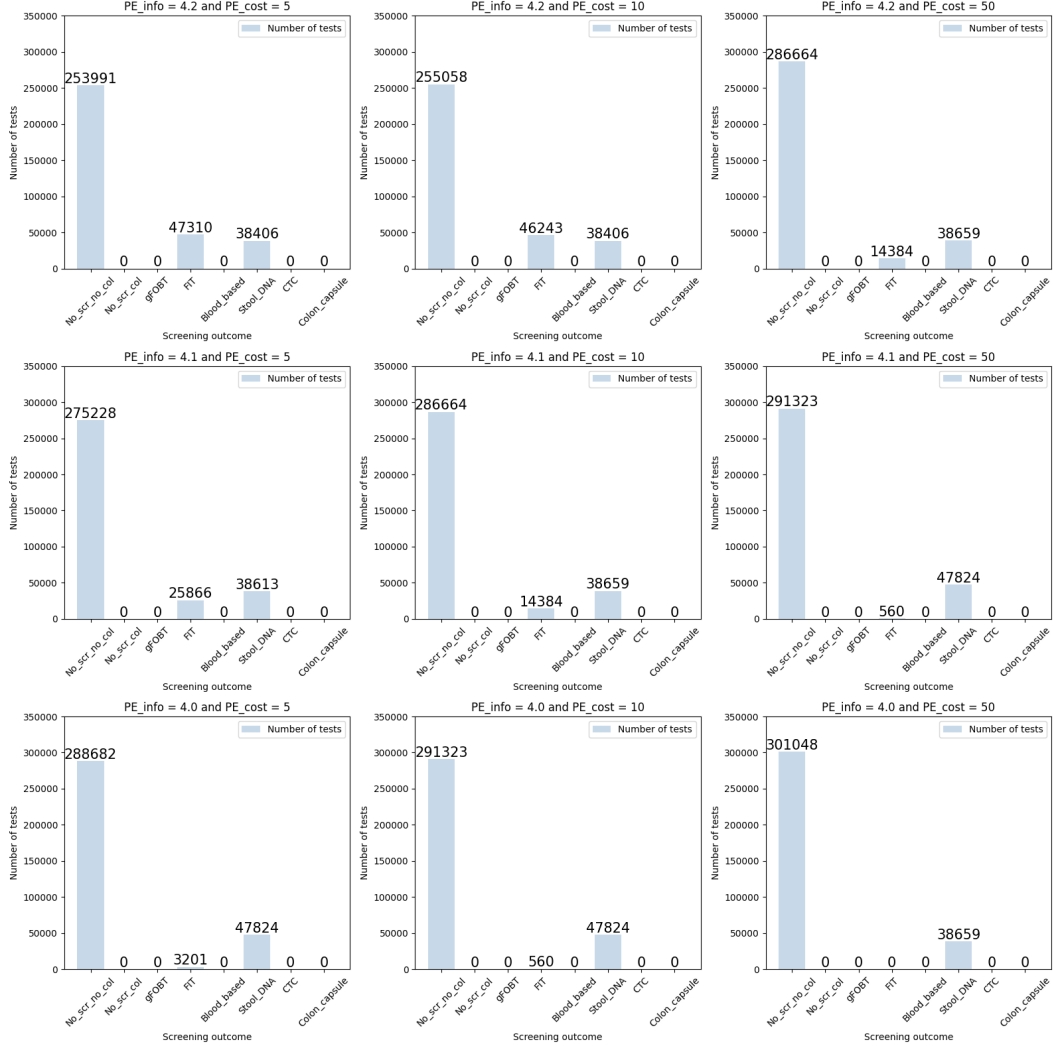

Figure 6: Plot with different PE info and PE cost

3.5.2 Sensitivity analysis with respect to 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. 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.

Figure 7 shows two screening test recommendation distributions. 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.
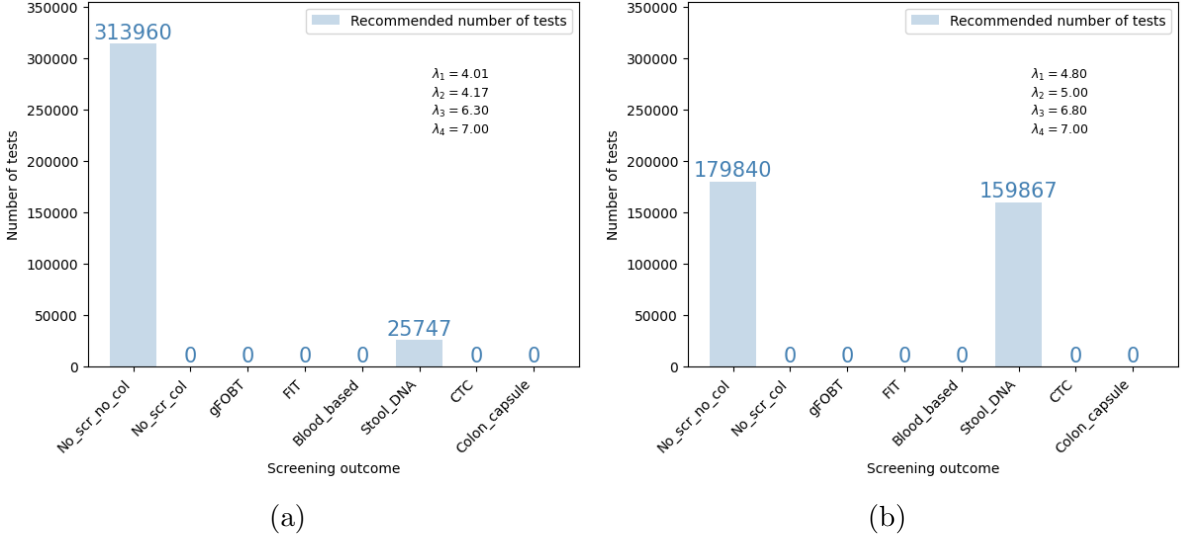

Figure 7: Screening count distribution for different λ values

Therefore, a moderate modification of comfort parameters impacts program results. To ensure consistency, we would recommend following the elicitation protocol in a moderate number of patients chosen at random, 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 decision support system for personalised screening in connection with CRC detection. Stemming from an earlier predictive model [14] we incorporated new chance, decision, and value 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 to strengthen the

monitoring systems of colorectal screening programmes, as has been proposed for European countries [29]. As showcased, this facilitates a more personalized approach to CRC screening 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 is the one discussed in Section 2.2.2, and assimilated to that of the policy makers designing or assessing a screening strategy. However, the preference model may be easily modified, and eventually individualised, to promote a more personalized approach. in line with [30].

As illustrated in Section 3.1, 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 and genetics is 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 personalising screening programmes. In any case, the major strength of this work lies in its development of a practical framework and methodology which is adaptable to more complex scenarios.

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 ONCO-SCREEN project with the methodology specified in Section 3.4. The second one, given the low acceptance of CRC screening programs in many countries, will focus on amplifying our model with incentive mechanisms to promote screening adoption.

References

[1] F. Kastrinos, S. S. Kupfer, and S. Gupta. Colorectal cancer risk assessment and precision approaches to screening: Brave new world or worlds apart? *Gastroenterology*, 164(5):812–827, 2023. ISSN 0016-5085. doi: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, and G. Van Hal. Towards novel non-invasive colorectal cancer screening methods: a comprehensive review. *Cancers*, 13(8):1820, 2021. doi:10.3390/cancers13081820.

[3] I. Ola, R. Cardoso, M. Hoffmeister, and H. Brenner. Utilization of colorectal cancer screening tests across european countries: a cross-sectional analysis of the european health interview survey 2018–2020. *The Lancet Regional Health–Europe*, 41, 2024. doi:10.1016/j.lanepe.2024.100920.

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

[5] S. Barré, H. Leleu, R. Benamouzig, J.-C. Saurin, A. Vimont, S. Taleb, and F. De Bels. Cost-effectiveness analysis of alternative colon cancer screening strate-

gies in the context of the french national screening program. *Therapeutic Advances in Gastroenterology*, 13:1756284820953364, 2020. doi: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): 51–59, 2013. doi:10.1055/s-0032-1325997.

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

[8] A. Fouladi, A. Asadi, E. A. Sherer, and M. Madadi. Cost-effectiveness analysis of colorectal cancer screening strategies using active learning and monte carlo simulation. *Medical Decision Making*, 44(5):554–571, 2024. doi:10.1177/0272989X241258224.

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

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

[11] R. D. Shachter. Evaluating influence diagrams. *Operations Research*, 34(6):871–882, 1986. doi:10.1287/opre.34.6.871.

[12] M. Roberts, L. B. Russell, A. D. Paltiel, M. Chambers, P. McEwan, and M. Krahn. Conceptualizing a model: a report of the ispor-smdm modeling good research practices task force–2. *Medical Decision Making*, 32(5):678–689, 2012. doi:10.1177/0272989X12454941.

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

[14] D. Corrales, A. Santos-Lozano, S. López-Ortiz, A. Lucia, and D. R. Insua. Colorectal cancer risk mapping through bayesian networks. *Computer Methods and Programs in Biomedicine*, page 108407, 2024. doi:10.1016/j.cmpb.2024.108407.

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

[16] D. Corrales. Decision_Model_Screening_CRC, May 2024. URL https://github.com/DanielCorralesAlonso/Decision_Model_Screening_CRC.

[17] S. French and D. Rios Insua. *Statistical Decision Theory*. Wiley, 2000.

[18] 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, and 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*, 34:59–76, 2016. doi:10.1007/s40273-015-0328-1.

[19] R. L. Keeney and R. S. Gregory. Selecting attributes to measure the achievement of objectives. *Operations Research*, 53(1):1–11, 2005. doi:10.1287/opre.1040.0158.

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

[21] T. M. Cover. *Elements of Information Theory*. John Wiley & Sons, 1999. doi:10.1002/047174882X.

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

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

[24] J. González-Ortega, V. Radovic, and D. Ríos Insua. Utility elicitation. *Elicitation: The science and art of structuring judgement*, pages 241–264, 2018. doi:10.1007/978-3-319-65052-4_10.

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

[26] P. H. Farquhar. State of the art—utility assessment methods. *Management Science*, 30(11):1283–1300, 1984. doi:10.1287/mnsc.30.11.1283.

[27] A. Pellat, J. Deyra, R. Coriat, and S. Chaussade. Results of the national organised colorectal cancer screening program with fit in paris. *Scientific reports*, 8(1):4162, 2018. doi:10.1038/s41598-018-22481-9.

[28] C. J. Van Rijsbergen. Information retrieval. 2nd. newton, ma, 1979.

[29] E. Altobelli, L. Rapacchietta, C. Marziliano, G. Campagna, V. F. Profeta, and R. Fagnano. Differences in colorectal cancer surveillance epidemiology and screening in the who european region. *Oncology Letters*, 17(2):2531–2542, 2019. doi:10.3892/ol.2018.9851.

[30] R. J. Volk, V. B. Leal, L. E. Jacobs, A. M. Wolf, D. D. Brooks, R. C. Wender, and R. A. Smith. From guideline to practice: new shared decision-making tools for colorectal cancer screening from the american cancer society. *CA: a cancer journal for clinicians*, 68(4):246–249, 2018. doi:10.3322/caac.21459.

A   Probabilities available from public sources

To assess the probabilities in the chance nodes introduced, Table 19 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.

	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

Table 19: Specificity, sensitivity, and cost in € of interventions.

We also require the probabilities and expected costs of complications as reflected in Table 20 based on [5].

	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 20: Probabilities and expected costs of complications for CRC interventions.

B   Full elicitation of comfort parameters

Table 21 contains full details of the calculations used to assess comfort parameters.

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€	$\lambda_3 = 5.04$	
	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€		$\bar{\lambda}_3 = 6.80$
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	–	–	–	$\lambda_4 = 7$

Table 21: Eliciting parameter λ_k. X indicates the preferred alternative in the corresponding pairwise comparison.