European Journal of Operational Research 238 (2014) 313–326

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Innovative Applications of O.R.

Using a partially observable Markov chain model to assess colonoscopy screening strategies – A cohort study

Y. Li [a], M. Zhu [a], R. Klein [b], N. Kong [a,*]

[a] Weldon School of Biomedical Engineering, Purdue University, 206 S. Martin Jischke Dr., West Lafayette, IN 47907, USA
[b] Medical Decision Modeling Inc., 8909 Purdue Road #550, Indianapolis, IN 46268, USA

ARTICLE INFO

Article history:
Received 18 October 2012
Accepted 1 March 2014
Available online 14 March 2014

Keywords:
Medical decision making
Cancer screening
Colorectal cancer natural history
Partially observable Markov chain
Cost-effectiveness analysis

ABSTRACT

Colorectal cancer (CRC) is notoriously hard to combat for its high incidence and mortality rates. However, with improved screening technology and better understanding of disease pathways, CRC is more likely to be detected at early stage and thus more likely to be cured. Among the available screening methods, colonoscopy is most commonly used in the U.S. because of its capability of visualizing the entire colon and removing the polyps it detected. The current national guideline for colonoscopy screening recommends an observation-based screening strategy. Nevertheless, there is scant research studying the cost-effectiveness of the recommended observation-based strategy and its variants. In this paper, we describe a partially observable Markov chain (POMC) model which allows us to assess the cost-effectiveness of both fixed-interval and observation-based colonoscopy screening strategies. In our model, we consider detailed adenomatous polyp states and estimate state transition probabilities based on longitudinal clinical data from a specific population cohort. We conduct a comprehensive numerical study which investigates several key factors in screening strategy design, including screening frequency, initial screening age, screening end age, and screening compliance rate. We also conduct sensitivity analyses on the cost and quality of life parameters. Our numerical result demonstrates the usability of our model in assessing colonoscopy screening strategies with consideration of partial observation of true health states. This research facilitates future design of better colonoscopy screening strategies.

Introduction

Colorectal cancer (CRC) ranks third in incidence among cancer diseases and second in cancer-related death in the U.S. (Jemal, Bray, & Center, 2011). Nonetheless, CRC is often cured if detected early, e.g., the 5-year survival rate for localized CRC is 90% while the survival rate is only 12% if the cancer has spread to distant locations (Howlader et al., 2012). However, there are often no symptoms when CRC is in its early stages. Fortunately, with improved fiber optic technology, enhanced understanding of CRC natural history, and more intelligent screening strategies, it is increasingly possible to detect polyps including precancerous adenomas early, predict their progression accurately, and thus reduce CRC incidence and mortality. Furthermore, slow precancerous adenoma progression allows intelligent application of screening to detect and remove adenomas before they become cancerous.

Colonoscopy is the most accurate CRC screening test as it provides a visual diagnosis of the entire colon and rectum. It can detect precancerous adenomas and remove them immediately before they become cancerous. This adenoma removal procedure, also called polypectomy, can significantly reduce patients' cancer risks. The American College of Gastroenterology (ACG) recommendations imply that colonoscopy is the preferred cancer screening method (Rex et al., 2009). The ACG further suggests that other cancer detection tests are less preferred but should be offered to patients who decline colonoscopy. In addition to colonoscopy, several CRC screening methods are currently used in practice. These include fecal testing for occult blood (i.e., g-FOBT, FIT, and i-FOBT), fecal DNA testing, flexible sigmoidoscopy, and computed tomographic colonography (virtual colonoscopy). Guidelines from the U.S. Multisociety Task Force (Levin et al., 2008) and the U.S. Preventive Services Task Force (U.S. Preventive Services Task Force, 2008) recommend some of the above alternative screening methods that are less invasive and less expensive than colonoscopy. However, these methods are only good at detecting preclinical cancer but not adenomas. Hence, their value is only significant in low risk populations. Therefore, our main objective is assessing colonoscopy screening strategies using an innovative mathematical model.

We consider two classes of colonoscopy screening strategies: fixed-interval screening strategy and observation-based screening

* Corresponding author. Tel.: +1 765 496 2467.
E-mail addresses: li528@purdue.edu (Y. Li), mzhu@purdue.edu (M. Zhu), rwk@mdm-inc.com (R. Klein), nkong@purdue.edu (N. Kong).

http://dx.doi.org/10.1016/j.ejor.2014.03.004
0377-2217/

strategy. With a fixed-interval screening strategy, patients are recommended to take the screening tests in a fixed time interval regardless of their cancer risks. An observation-based screening strategy, however, specifies the timing of the next screening based on the previous screening result. Intuitively, a well designed observation-based screening strategy should be more desirable than a fixed-interval screening strategy since it determines screening intervals based on an individual's cancer risk rather than treating all patients the same. Thus, designing a good observation-based screening strategy is an important research question.

We develop a discrete-time partially observable Markov chain (POMC) model with a detailed description of precancerous adenoma states and a set of age-dependent transition probabilities estimated from a large longitudinal clinical data set for a specific population cohort. Traditionally, Markov models are used to represent the transitions among the true adenoma states. However, the true adenoma states can rarely be observed with complete accuracy due to limitations of the technology and insufficient experience of the practitioner who performs the test. For example, based on Rex et al. (1997), only about 70% of small adenomas (size less than 5 mm) are detected by a single colonoscopy. Thus, we use belief states to capture the likelihood of each true state being occupied. We update the belief states in a Bayesian manner based on the latest colonoscopy findings and the disease progression, i.e., natural history. By using a detailed description of precancerous adenoma states, the state space of our POMC model becomes much larger compared to the existing Markov models in the literature. Furthermore, by incorporating incomplete adenoma detection and removal, optimization of screening strategies with the POMC model becomes extremely challenging computationally. Therefore, we focus on assessing the cost-effectiveness of the screening strategies and investigating the effects of several key factors in the strategy design, including screening frequency, initial screening age, screening end age, and partial compliance to screening tests.

Our work is among the first that applies POMC modeling to assess colonoscopy screening strategies. Our main contributions are twofold. First, we incorporate inaccurate observations of health states and update the belief state based on the colonoscopy test results in a Bayesian manner. Such incorporation of partial observability has not been seen in the literature of economic analysis for CRC screening. Second, we conduct comprehensive cost-effectiveness assessment and compare fixed-interval and observation-based colonoscopy screening strategies. Although our results may lack generalization because our parameter estimations are based on a specific population cohort, the modeling framework is valuable and can be easily adapted to assess colonoscopy screening strategies for any other cohort once its clinical data is available.

The remainder of this paper is organized as follows. In section 'Literature review', we provide literature review on both well-accepted and recent economic studies and decision models on CRC screening strategy design. In section 'Model development', we present our POMC model and describe the belief update and outcome measures. In section 'Parameter estimation and experiment design', we describe our data sources, parameter estimation, and experimental design. We report numerical studies with a baseline case study and several sensitivity analyses in section 'Numerical results'. Conclusions and future research directions are presented in section 'Conclusions and future work'.

Literature review

Long duration of CRC progression at the precancerous stages and availability of various screening methods motivate the development of accurate CRC disease models and the analysis of cost-effectiveness for CRC screening. Pignone et al. (2005) and Zauber et al. (2012) summarized most existing CRC models which can be divided into two categories: discrete-event based models and Markov based models. Discrete-event simulation models (Cubbage, 2004; Ness, Holmes, Klein, & Dittus, 2000; Roberts, Wang, Klein, Ness, & Dittus, 2007; Tafazzoli, Roberts, Ness, Klein, & Dittus, 2009; Loeve, Boer, van Oortmarssen, van Ballegooijen, & Habbema, 1999; Loeve et al., 2000; Rutter, Zaslavsky, & Feuer, 2010; Wilschut et al., 2011) simulate a population of individuals from birth to death. Each simulated individual experiences a series of events, including colorectal adenoma incidence, growth, and transition, CRC staging, CRC or non-CRC induced deaths, CRC screening tests, and adenoma removals. The cost and effectiveness outcomes can be obtained via the simulation. Discrete-event simulation models suffer from complexity that hinders transparency as well as the need of extensive data for calibration. Markov based models (Frazier, Colditz, Fuchs, & Kuntz, 2000; Sonnenberg, Delco, & Inadomi, 2000; Vijan, Hwang, Hofer, & Hayward, 2001; Song, Fendrick, & Ladabaum, 2004; Ladabaum, Song, & Fendrick, 2004; Heitman, Hilsden, Au, Dowden, & Manns, 2010; van Rossum et al., 2011; Sobhani, Alzahouri, Ghout, Charles, & Durand-Zaleski, 2011; Hedden et al., 2012; Lucidarme et al., 2012), on the other hand, specify CRC-related health states individuals may occupy during their lifetimes and use the Markovian property to guide state transitions in a discrete fashion. The occurrence of CRC screening alters the state transition from natural disease progression. These Markov chain-based models differ in CRC-related health states definition, state transition probability, time horizon, and outcome parameters, but all of them assume the health states of a person are explicitly observed, which is not necessarily valid due to the asymptomatic nature of early-stage CRC. It is worth noting that three models have been approved by the Cancer Intervention and Surveillance Modeling Network (CISNET), which represents the state of the art for the CRC screening models (National Cancer Institute, 2012). They are MISCAN-Colon model (Loeve et al., 1999; Loeve et al., 2000), Sim-CRC model (Frazier et al., 2000), and CRC-SPIN model (Rutter & Savarino, 2010).

Through literature review, we identify two important issues that, to the best of our knowledge, are not fully addressed in the existing CRC screening strategy assessment literature. The first issue is the assessment of observation-based colonoscopy screening strategies. Even though the current guideline developed by the U.S. Preventive Work Force has recommended the observation-based strategies for patients with different colonoscopy screening results in terms of the number of precancerous adenomas and the size of each of the adenomas (Levin et al., 2008), we have not witnessed any study that assesses the cost-effectiveness of the observation-based strategies and their variants. To achieve this, a more detailed description of the CRC natural history model is required, which implies an expansion of the state space in the existing Markov based models and requires more complicated model calibration with detailed colonoscopy observation data. In addition, adenoma removal via polypectomy is a unique feature associated with colonoscopy, which requires the incorporation of reverse transitions in the model. The other issue is the partial observability of patients' health states from colonoscopy screening tests. Most previous Markov-based cancer screening models assume patients' health conditions can be fully observed, which is not realistic. Maillart, Ivy, Ransom, and Diehl (2008) and Ayer, Alagoz, and Stout (2012) incorporated partial observability in the design of mammography screening strategy, and Zhang, Denton, Balasubramanian, Shah, and Inman (2012a, 2012b) proposed partially observable Markov decision process (POMDP) models to study prostate cancer screening decision making. All of these models either for breast cancer or prostate

cancer capture the precancerous stage using one aggregate state or a few states in terms of cancer risk. However, the unique mechanism of CRC progression motivates modelers to capture the precancerous stage with a much larger state set based on the number and size of precancerous adenomas, which subsequently makes incorporation of partial observability more difficult. We describe the development of a POMC model to address these two issues.

Model development

We develop a finite-horizon discrete-time partially observable Markov chain (POMC) model to capture the incidence and progression of precancerous adenomas and CRC (Fig. 1), and the effect of colonoscopy screening with polypectomy (Fig. 2). In our POMC model, patients transition through a series of precancerous adenoma, preclinical cancer, clinical cancer, and death states. A colonoscopy test helps assess patients' CRC-related health states and may subsequently lead to removal of detected adenomas, so patients' disease states can regress and their follow-up tests may differ. However, colonoscopy is not entirely accurate in detecting adenomas and thus a patient's disease state is only *partially* observed. We use a probability distribution over all the states to represent the *belief* on the true disease state and update the belief in a Bayesian manner based on the colonoscopy screening result at each discrete time point. We assign a quality of life multiplier to each disease state and specify the costs of colonoscopy screening and CRC treatment. For a given screening strategy, we calculate the expected cumulative quality-adjusted life years (QALYs) and the expected cumulative cost over the studied period. We next present the detailed model formulation.

Time horizon

Our model has a time horizon $t = \{40, 41, 42, \ldots, 100\}$. In the model, a person's disease state evolves from age 40 to 100. We assume that a screening colonoscopy can only be performed at the beginning of each year. So the time interval between state transitions is 1 year. In the following exposition, we define the states and observations at discrete age points.

States

The states in our model are associated with precancerous adenoma and CRC incidence and progression as well as patient mortality. As shown in Fig. 1, there are an adenoma-free state (AF); an aggregated state set (PCA) which contains 83 distinct precancerous adenoma states; three preclinical (asymptomatic) cancer states, i.e., local (PL), regional (PR), and distant (PD); three clinical (symptomatic) cancer states (CL, CR, CD), corresponding to the three preclinical cancer states; and two states for death either induced by CRC ($D1$) or due to other reasons ($D2$). Thus, the state space is defined as $S = \{AF, PCA_{(1)}, PCA_{(2)}, \ldots, PCA_{(83)}, PL, PR, PD, CL, CR, CD, D1, D2\}$.

Our state space is an expansion from those commonly used in the recent literature. That is, instead of having only a few precancerous states based on number of adenomas and size or histology of the adenomas (Zauber et al., 2008), we model each precancerous adenoma with an independent adenoma-carcinoma sequence and capture the concurrent progression of the adenomas. Thus, the incidence of an adenoma is independent of the number of adenomas already presented and the progression of each adenoma is independent of the progression of other existing adenomas. We use a triple (n_s, n_m, n_l) to denote each adenoma state, where n_s, n_m, and n_l denote the numbers of small ($\leqslant$5 mm), medium

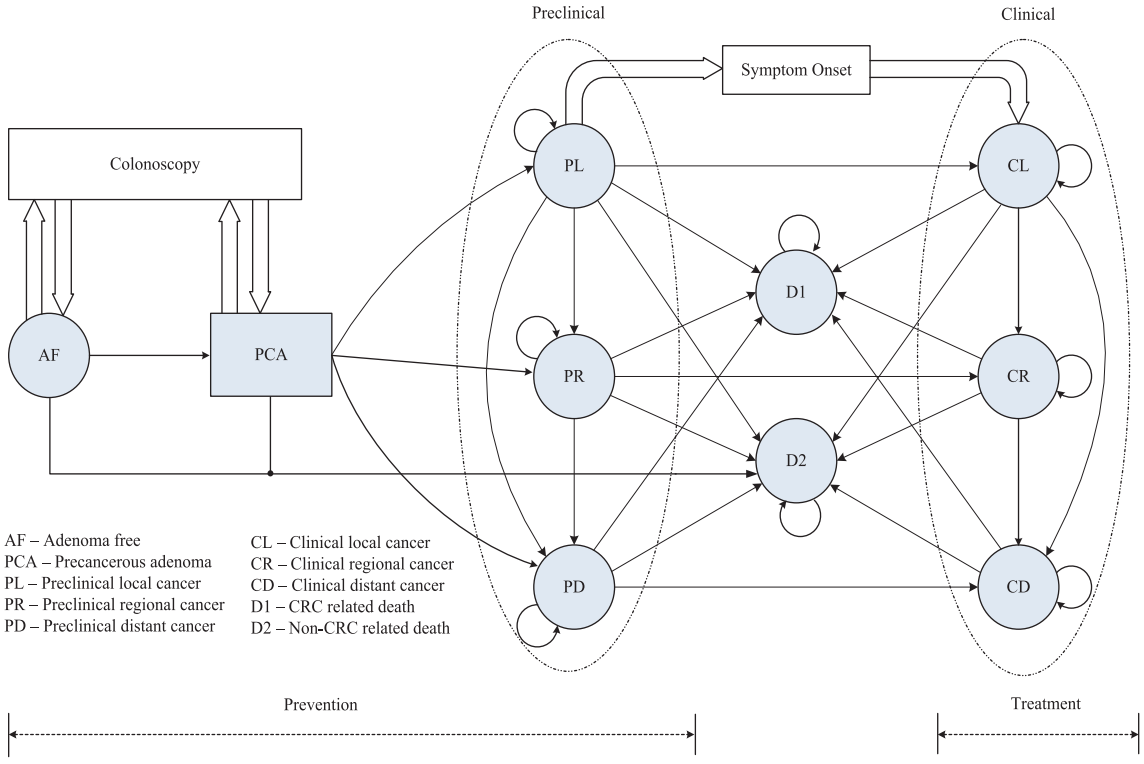

Fig. 1. State transition diagram.

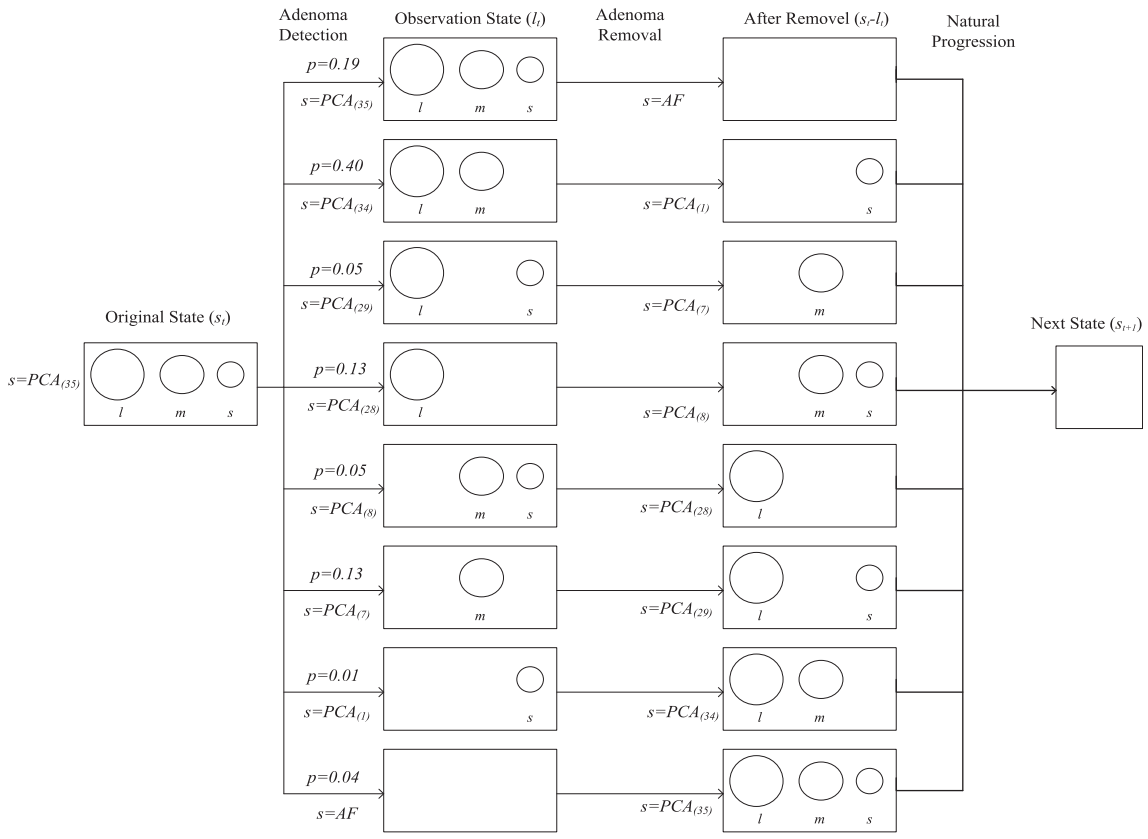

Fig. 2. Illustration of precancerous state transition after colonoscopy.

(5–10 mm), and large ($\geqslant$10 mm) adenomas, respectively. Although the location of adenomas may affect the sensitivity of colonoscopy for detecting adenomas in practice, we do not consider the location in the adenoma state description. Because fewer than 5% of colonoscopy tests detect more than six adenomas (Sherer, Imperiale, Ambedkar, Perng, & Yih, 2010), we limit the total number of adenomas to six (i.e., $n_s + n_m + n_l \leqslant 6$). So a precancerous adenoma state is a combination of small, medium, and large adenomas with a total number of at most six, and the total number of possible combinations is 83 (see Appendix A for detailed definitions of the 83 precancerous adenoma states). In addition, the preclinical cancer means that the cancer is neither symptomatic nor diagnosed. In a preclinical stage, the cancer is present in the body but unknown to the patient and her attending physician. In a clinical stage, a cancer is either symptomatic or diagnosed. In our model, when a patient is in a preclinical cancer state, she will transfer to the corresponding clinical cancer state with probability 1 when there is cancer-related symptom onset or a scheduled colonoscopy screening is performed. The CRC staging system (i.e., local, regional, distant cancer) used in our model, which considers stage and extent of the cancer, has been used in several CRC natural history models (see e.g., Frazier et al., 2000; Ladabaum et al., 2004; Vijan et al., 2001; Ness et al., 2000). An alternative staging system, consisting of four stages, has also been used in the recent literature (see e.g., Zauber et al., 2008; Lansdorp-Vogelaar et al., 2009).

Observations

A person's health state cannot be observed without a colonoscopy test unless she has progressed to a clinical cancer state or death. Even with colonoscopy, a person's health state cannot be fully observed due to the technological limitations in colonoscopy. However, a person could obtain some information on her specific health state and could thus make further decisions based on the observation. We use *NO* to denote no observation when no colonoscopy is performed to an asymptomatic person; AF^o to denote the observed adenoma-free state; $PCA^o_{(i)}$, $i = 1, \ldots, 83$, to denote the observed 83 precancerous adenoma states; *CA* to denote CRC observed either through colonoscopy or onset of symptoms; and *D* to denote death. Thus the observation space is defined as $O = \{NO, AF^o, PCA^o_{(1)}, PCA^o_{(2)}, \ldots, PCA^o_{(83)}, CA, D\}$.

Actions

The possible actions at the beginning of each time interval are performing colonoscopy test (*C*) or waiting until next time (*W*). We denote the action taken for a person at age t by a_t, thus $a_t \in A = \{C, W\}$ for all ages t.

Observation probabilities

The observation at each discrete age point is dependent on the underlying true disease state and the action taken, and it can be probabilistically represented by the state/action pair. We use l_t to denote the observation made to a person at age t and use $q_t(l_t \mid s_t, a_t)$ to denote the probability of observing $l_t \in O$ if the person is in true state $s_t \in S$ and takes action $a_t \in A$ at age t. The observation probability can be estimated for different true states $s_t \in S$ as follows.

- $s_t \in \{AF, PCA_{(1)}, PCA_{(2)}, \ldots, PCA_{(83)}\}$. If $a_t = W$, we have no information regarding the person's health state, so we have $q_t(NO \mid s_t, W) = 1$ and $q_t(l_t \mid s_t, W) = 0$ for $l_t \in O \setminus \{NO\}$. If $a_t = C$, since colonoscopy is not completely accurate, the

observation probability is estimated based on the sensitivity of detecting adenomas. We let p_s^{test}, p_m^{test}, p_l^{test} denote the probabilities with which a small, medium, and large sized adenoma is detected and removed, respectively. Then the probability of observing l_t given true state s_t is

$$q_t(l_t \mid s_t, C) = \binom{n_s^{s_t}}{n_s^{l_t}} (1-p_s^{test})^{n_s^{s_t}-n_s^{l_t}} (p_s^{test})^{n_s^{l_t}} \times \binom{n_m^{s_t}}{n_m^{l_t}} \times (1-p_m^{test})^{n_m^{s_t}-n_m^{l_t}} (p_m^{test})^{n_m^{l_t}} \times \binom{n_l^{s_t}}{n_l^{l_t}} (1-p_l^{test})^{n_l^{s_t}-n_l^{l_t}} \times (p_l^{test})^{n_l^{l_t}}, \quad l_t \in O \setminus \{NO, CA, D\}, \tag{1}$$

and $q_t(l_t \mid s_t, C) = 0$ for $l_t \in \{NO, CA, D\}$.

- $s_t \in \{PL, PR, PD\}$. If $a_t = W$, since no symptoms are observed in each preclinical cancer state, we do not schedule any colonoscopy, i.e., $q_t(NO \mid s_t, W) = 1$ and $q_t(l_t \mid s_t, W) = 0$ for $l_t \in O \setminus \{NO\}$. If $a_t = C$, the colonoscopy serves as a diagnostic tool and we assume that CRC can be identified with certainty, i.e., $q_t(CA \mid s_t, C) = 1$ and $q_t(l_t \mid s_t, C) = 0$ for $l_t \in O \setminus \{CA\}$.
- $s_t \in \{CL, CR, CD\}$. When a patient's disease condition has progressed to a clinical cancer state, it implies that the cancer has been diagnosed and thus no colonoscopy is needed and the observation on the state is deterministic and reflects the true state, i.e., $q_t(CA \mid s_t, W) = 1$ and $q_t(l_t \mid s_t, W) = 0$ for $l_t \in O \setminus \{CA\}$.
- $s_t \in \{D1, D2\}$. If a patient is dead, no colonoscopy is needed, and the only observation is death, i.e., $q_t(D \mid s_t, W) = 1$ and $q_t(l_t \mid s_t, W) = 0$ for $l_t \in O \setminus \{D\}$.

Transition probabilities

We use $p_t(s_{t+1} \mid s_t, a_t)$ to denote the state transition probability from disease state s_t to s_{t+1} given action a_t at age point t. If $a_t = W$, a person's CRC-related health state evolves naturally. We next present several specifications on the state transitions.

- *Irreversible progression*. We assume that neither the number of adenomas nor the size of each adenoma decreases unless a colonoscopy with polypectomy is performed. In addition, we assume that a patient at a preclinical cancer state cannot return to any precancerous adenoma state or the adenoma-free state, and the transitions among preclinical cancer states are irreversible as well. Finally, we assume that a patient in a clinical cancer state can only remain in the same state or transition to one of the two death states regardless of the action taken.
- *Symptom development*. We assume that a person does not develop any CRC-related symptoms in the adenoma-free state or any precancerous adenoma states, so CRC treatments are not received in those states. At a preclinical cancer state, a person may develop CRC-related symptoms, and we assume that she immediately takes a test for the diagnostic purpose. We also assume that such a person with CRC-related symptoms is diagnosed with certainty, which implies the transition from the preclinical cancer state to the corresponding clinical cancer state with the onset of the symptoms.

If $a_t = C$, the colonoscopy test changes the course of CRC natural disease progression in that it either regresses the disease at a precancerous adenoma state or triggers treatment at a preclinical cancer state.

- For $s_t \in \{AF, PCA_{(1)}, PCA_{(2)}, \ldots, PCA_{(83)}\}$, we incorporate the screening result to calculate the corresponding probability of each possible realization of state s_{t+1}. Fig. 2 shows an example of precancerous adenoma state transition after a colonoscopy. Suppose a patient is at state $PCA_{(35)}$ at time t, which means she has one large, one medium, and one small adenoma. Due to the limitation of the test, it is possible that only a portion of the adenomas will be detected through colonoscopy. As the figure shows, there is only 19% of chance that all the adenomas can be detected, but there is 40% of chance that the small adenoma is missed in the test (corresponding to the state $PCA_{(34)}$). Since colonoscopy can only remove the adenomas it detected, there is only a 19% chance that the patient will be adenoma free after adenoma removal, while there is a 40% chance that the patient will be at state $PCA_{(1)}$. The figure further shows, for a patient at state $PCA_{(35)}$, there are eight possible states with different probabilities after a colonoscopy is performed. These possible health states will still follow natural history progression for 1 year and form the belief of state at time $(t+1)$.
- For $s_t \in \{PL, PR, PD\}$, we assume that a screening test diagnoses CRC and identifies its state with certainty. Then a preclinical state transitions to the corresponding clinical state with certainty, i.e., $p_t(CL \mid PL, C) = 1, p_t(CR \mid PR, C) = 1$, and $p_t(CD \mid PD, C) = 1$, and the probability of transitioning to any other state is 0.
- For $s_t \in \{CL, CR, CD\}$, we assume that any colonoscopies after local or regional cancer are part of the standard of care and patients in clinical cancer states transition to more advanced cancer states or death states based on transition probabilities estimated from Sherer et al. (2010) and Sherer, Imler, and Imperiale (2012).
- For $s_t \in \{D1, D2\}$, no colonoscopy is needed.

Belief update

In our POMC model, we update the belief on the CRC-related disease states at each age point in a Bayesian manner. We denote a belief state to be $\pi_t = (\pi_t(AF), \pi_t(PCA_{(1)}), \pi_t(PCA_{(2)}), \ldots, \pi_t(PCA_{(83)}), \pi_t(PL), \pi_t(PR), \pi_t(PD), \pi_t(D1), \pi_t(D2), \pi_t(CL), \pi_t(CR), \pi_t(CD))$, where $\pi_t(i)$ for $i \in S$ is the probability with which a person is in state i at age t. We let Π be the set containing all possible belief states, i.e., $\pi_t \in \Pi \equiv \{\pi_t \in \mathbb{R}^{92} \mid \sum_{i \in S} \pi_t(i) = 1, \pi_t(i) \geqslant 0, i \in S\}$.

We next describe how to update the belief state from one age point to the next. At any age point, if $a_t = W$, we update the belief state π_t following the CRC natural history, i.e.,

$$\pi_{t+1}(s_{t+1}) = \sum_{s_t \in S} \pi_t(s_t) p_t(s_{t+1} \mid s_t, W). \tag{2}$$

If $a_t = C$, we update the belief state π_t based on the observation l_t. If $l_t \in \{PCA^o_{(1)}, PCA^o_{(2)}, \ldots, PCA^o_{(83)}\}$, colonoscopy may not accurately identify each adenoma and thus may not remove all the adenomas. In this case, we use the following three steps to update the belief state.

1. *Bayesian update*. We first apply Bayes' theorem to update the belief state based on the screening test result. Recall $q_t(l_t \mid s_t, C)$ is the probability of observing state l_t with colonoscopy, given the true state s_t. We let $\pi'_t(\cdot)$ be the updated belief after step 1. Then for the true state s_t, we have

$$\pi'_t(s_t) = \frac{q_t(l_t \mid s_t, C)\pi_t(s_t)}{\sum_{s_t \in S} q_t(l_t \mid s_t, C)\pi_t(s_t)}. \tag{3}$$

2. *Adenoma removal*. We then consider the dynamics caused by adenoma removals. We assume that all adenomas can be removed as long as they are detected. For a person in state s_t, i.e., $(n_s^{s_t}, n_m^{s_t}, n_l^{s_t})$, if the observation is l_t, i.e., $(n_s^{l_t}, n_m^{l_t}, n_l^{l_t})$, the state after adenoma removal is $(n_s^{s_t} - n_s^{l_t}, n_m^{s_t} - n_m^{l_t}, n_l^{s_t} - n_l^{l_t})$ (or $s_t - l_t$

with shorthand notation). We let $\pi''_t(\cdot)$ be the updated belief after step 2, and $s'_t = s_t - l_t$, then we have

$$\pi''_t(s'_t) = \pi'_t(s_t). \quad (4)$$

3. *Natural history progression*. After the above two steps, the effects of interventions on the belief update have been addressed. We finally update the belief state based on the CRC natural history to obtain

$$\pi_{t+1}(s_{t+1}) = \sum_{s'_t \in S} \pi''_t(s'_t) p_t(s_{t+1} \mid s'_t, W). \quad (5)$$

If $l_t = AF^o$, i.e., no adenomas are found, the belief update only involves steps 1 and 3 described above. If $l_t = CA$, no belief needs to be updated since the true disease state is observable. Hence, the person transitions to the corresponding clinical cancer state with certainty.

Outcome measures

We use cumulative quality-adjusted life years (QALYs) and cumulative CRC-related cost as the outcome measures. QALY is an effectiveness measurement which quantifies the yearly-specific health utilities of a person given a certain disease state. Let $u_t(s_t)$ be the yearly utility value (between 0 and 1) for a person of age t in state s_t. Then cumulative QALYs are calculated as $\sum_{t=t_0} e^{-\lambda_E(t-t_0)} u_t(s_t)$, where t_0 is the starting age of screening strategy evaluation (i.e., in our numerical studies, it is 40), and λ_E is a discount factor between 0 and 1. The cumulative CRC-related cost consists of both colonoscopy screening cost and cancer treatment cost. In addition, it may include the cost incurred by polypectomy if any adenomas are detected during the screening test. We let c_{col} denote the cost for performing a colonoscopy test and c_{pol} denote the cost for performing a colonoscopy test and subsequent polypectomy. Depending on the stage of cancer treatment, the treatment cost can be further divided into initial treatment cost, denoted by c_{ini}, yearly variable cost, denoted by c_{var}, and terminal cost, denoted by c_{ter}. Hence, the cost for each individual at age t, denoted by $c_{cum}(s_t, a_t)$, is calculated as $c_{col}\gamma_{col}(s_t, a_t) + c_{pol}\gamma_{pol}(s_t, a_t) + c_{ini}\gamma_{ini}(s_t, a_t) + c_{var}\gamma_{var}(s_t, a_t) + c_{ter}\gamma_{ter}(s_t, a_t)$, where $\gamma_{col}(s_t, a_t)$, $\gamma_{pol}(s_t, a_t)$, $\gamma_{ini}(s_t, a_t)$, $\gamma_{var}(s_t, a_t)$, $\gamma_{ter}(s_t, a_t)$ are 0–1 variables indicating whether the cost incurred at age t is contributed by each of the five respective actions with (s_t, a_t). Then the cumulative CRC-related cost is calculated as $\sum_{t=t_0} e^{-\lambda_C(t-t_0)} c_{cum}(s_t, a_t)$, where λ_C is a discount factor between 0 and 1.

In some numerical studies, we also calculate the incremental cost-effectiveness ratio (ICER). ICER is a more comprehensive index to evaluate the tradeoff between effectiveness and cost. We use "no screening" as the baseline policy with which a person would not receive any colonoscopy screening through her lifetime. Then, ICER is interpreted as the cost required to gain one QALY with a given screening strategy relative to no screening. Let C_0, E_0, C, and E denote the cumulative CRC-related cost and QALYs under no screening and tested screening strategies, respectively. Then ICER for the tested screening strategy is calculated as $(C - C_0)/(E - E_0)$. For two strategies both with non-negative ICER values, the strategy with lower ICER is considered to be more cost-effective.

Parameter estimation and experiment design

Parameter estimation

Estimation of the state transition matrix is a large obstacle to applying a POMC model, especially for our detailed CRC natural history model with 92 states. To the best of our knowledge, there is no national clinical study which collects adenoma data from colonoscopy tests. Fortunately, we had access to the clinical data of about 4000 patients who received colonoscopy screening tests at the Roudebush Veterans Affairs Medical Center in Indiana, USA, from 2003–2009 (Sherer et al., 2012). This large longitudinal data set and the corresponding natural history model developed in Sherer et al. (2010) are sufficient to estimate the transition matrix in our POMC model and demonstrate the usability of the model. In particular, Sherer et al. (2010) provides us with age-dependent adenoma transition rates (i.e., the rates at which small adenomas appear and grow to larger adenomas with patient age) and adenoma–carcinoma transition rates (i.e., the rates at which adenomas with different sizes transition to carcinoma with respect to patient age). The researchers in Sherer et al. (2010) and Sherer et al. (2012) helped us validate our model with regard to the specific population cohort from which the clinical data was collected. It is worth noting that our cohort of patients has much higher risk of developing CRC than the general population nationwide. We demonstrated the cohort bias by comparing the age-dependent CRC incidence rates and cause-specific mortality rates obtained from the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) Program (Howlader et al., 2012) and the simulation results from our natural history model. The comparison is shown in Appendix B.

In addition, we acquired age-specific mortality risk for the general population, denoted by $p_t(D)$, from the Centers for Disease Control and Prevention (CDC) (Arias, 2010). Then we calculated the age-specific mortality risk due to other causes, denoted by $p_t(D2)$, by subtracting the CRC-induced mortality risk from the overall mortality risk, i.e., $p_t(D2) = p_t(D) - \sum_{s_t \in \{PL,PR,PD,CL,CR,CD\}} p_t(D1 \mid s_t)$. Note that CRC-induced mortality can only occur when the patient is either at a preclinical state or a clinical state. With $p_t(D2)$, we adjusted all transition probabilities by conditioning the transitions on survival from other death causes. Finally, we acquired the sensitivity of colonoscopy for detecting adenomas of different sizes from Rex et al. (1997), and acquired CRC-related utility coefficients and CRC screening and treatment cost coefficients from Tafazzoli et al. (2009). We summarize these model coefficients in Table 1.

Experiment design

There are two objectives in our numerical studies. One is to investigate the effects of several important factors related to CRC screening strategy design, including the age to initiate screening, the age to stop screening, and the screening compliance rate. The other is to investigate the sensitivity of our baseline conclusions to changes in the modeled cost and quality of life parameters. Once the values of aforementioned factors and model parameters were fixed, we constructed an experiment.

For each experiment, we assessed several observation-based colonoscopy screening strategies including the strategy currently recommended by the screening guideline. The current guideline recommends that people at average risk receive the initial colonoscopy test at age 50 and receive subsequent tests based on the results of prior tests (Levin et al., 2008). In general, colonoscopy results are clinically classified into four distinct groups in terms of precancerous adenoma prevalence: (1) no adenomas; (2) 1–2 small or medium (non-advanced) adenomas; (3) three or more small or medium adenomas; and (4) one or more large (advanced) adenomas. The current guideline suggests that the observation-based colonoscopy screening interval be 10 years if the test result is in group 1; 5–10 years if in group 2; and 3 years

Table 1
Model parameters, specific values and data sources.

Model parameter	Baseline value (min–max)	Data source
Transition among precancerous adenoma states and preclinical cancer states	Age specific	Sherer et al. (2010) and Sherer et al. (2012)
Mortality risk of the general population	Age specific	Arias (2010)
Sensitivity of colonoscopy (%)		Rex et al. (1997)
Small adenoma ($\leqslant$5 mm)	73	
Medium adenoma (5–10 mm)	87	
Large adenoma ($\geqslant$10 mm)	94	
CRC-related quality of life multipliers		Tafazzoli et al. (2009)
Adenoma-free	1	
Precancerous adenoma states	0.955 (0.91–1)	
Local cancer	0.61 (0.5–0.74)	
Regional cancer	0.605 (0.5–0.7)	
Distant cancer	0.25	
Death	0	
Cost of screening and treatment ($)		Tafazzoli et al. (2009)
Screening colonoscopy	614 (491–982)	
Colonoscopy with polypectomy	745 (596–1192)	
Initial treatment for local	20,323 (16,258–32,517)	
Initial treatment for regional	23,368 (18,694–37,389)	
Initial treatment for distant	26,708 (21,366–42,733)	
Yearly treatment for local	539 (431–862)	
Yearly treatment for regional	2461 (1969–3938)	
Yearly treatment for distant	26,855 (21,484–42,968)	
Terminal	21,172 (16,938–33,375)	

for groups 3 and 4 (Levin et al., 2008). Since the recommended guideline only gives a range on the screening interval if the current test result falls into group 2, we assessed both strategies with the interval at either lower or upper bound for this group. We term the strategy with the lower and upper bounds the *1st current* and *2nd current* strategies, respectively. We also assessed two alternative strategies that differ from the current guidelines in the screening interval length for certain observation group. We term the two alternatives the *conservative* and *aggressive* strategies, respectively. With the conservative strategy, a person is recommended to schedule her next test within 10 years if her current test result falls into group 3. With the aggressive strategy, a person is recommended to schedule her next test within 3 years if her current test result falls into group 2. As references, we also assessed the strategy that only allows one colonoscopy test throughout the lifetime, termed as the *one-time* strategy, and the strategies that fixed screening intervals at 10 years and 20 years regardless of the current test result, termed as the *routine 10* and *routine 20* strategies. Note that the one-time strategy has been investigated in Ness et al. (2000), which suggested such investigation is critical as determining the age of initial screening is key to screening strategy design. Also note that the fixed-interval screening strategies were also investigated by the CRC research community in recent years (see e.g., Sonnenberg et al., 2000; Tafazzoli et al., 2009).

We next investigated the effects of several parameters related to CRC screening strategy design, including the age to initiate screening, the age to stop screening, and the screening compliance rate. We conducted the following experiments: (1) varying the initial screening age to be 45, 50, and 55; (2) varying the screening end age to be 75, 80, and 85; and (3) varying the compliance rate to be 45%, 80%, and 100%. Finally, we conducted sensitivity analyses on the cost and quality of life coefficients.

Using the natural history model in Sherer et al. (2010), the same model embedded in our POMC model, we estimated the CRC-related risk distribution at age 40 and used it as the initial belief in our model. Note that the natural history model was developed for the entire lifetime and so it was run from the birth to age 40 for the estimation. Our estimate shows that without screening prior to age 40, any representative person from the average-risk population is believed to have 38% of chance being adenoma-free; 45% of chance having 1–2 small or medium adenomas; 8% of chance having 3 or more small or medium adenomas but no large adenomas; 7% of chance having at least one large adenomas; and 2% of chance having developed CRC. Note that with our model, one can vary the initial belief based on any real or hypothetical cohort. Given the main purpose of this paper being the presentation of our POMC model, we will analyze the effect of the initial belief in future research.

For each tested strategy, we computed the cumulative QALYs and CRC-related cost for an average-risk person from age 40 to her death. We considered a 3% discounting factor in our numerical studies, which is commonly used in cost-effectiveness analyses in health and medicine (Gold et al., 1996). We took a scenario tree enumeration approach to compute the expected QALYs and cost. The scenarios represent distinct CRC-related paths for the patient during her remaining lifetime. In other words, each scenario differs from others along the tree at some point based on the combination of CRC natural history dynamics and observation-based screening schedule. The occurrence probability of each scenario was computed by a sequence of probability multiplications along the scenario path. Once a scenario was realized, i.e., a leaf node of the scenario tree was reached, we computed its cumulative QALYs and cost as well as specified the occurrence probability of the scenario. Hence, we could compute the expectation by exhaustively enumerating all the scenarios. With a finite number of discrete age points, such computation could be done in finitely many steps. However, to reduce the computational burden, we used a probability threshold to control the number of scenarios to be enumerated. We ignored any scenarios with occurrence probability below the threshold. It is worth noting that there is no scenario that would lead to a substantially larger cost or quality of life than others. So a small occurrence probability would lead to a small contribution from the scenario to the expectation. Our objective was to yield reliable results within reasonable time. After some preliminary

tests, we gained experience on how the threshold affected the computational time and evaluation accuracy. For example, the number of enumerated scenarios would increase if the initial screening is performed at an earlier age or the screening interval is smaller. We eventually set the threshold to be 10^{-7}. When the initial screening age is 50, this threshold resulted in the numbers of enumerated scenarios to be 1.4×10^3, 1.8×10^5, 1.9×10^4, 5.6×10^5, 2.6×10^5, 2.0×10^5, and 7.8×10^5 for the one-time, routine 10, routine 20, 1st current, 2nd current, conservative, and aggressive strategies, respectively. All the evaluations were completed within one day.

Numerical results

Baseline analysis

In this section, we examine the cost-effectiveness of all the tested strategies using the baseline costs and quality of life parameters presented in Table 1 under baseline scenario. The baseline scenario is defined as a screening strategy with initial screening age 50, screening end age 80, and 100% compliance rate. We present the cumulative QALYs, cumulative costs, and ICERs versus no screening for all the tested screening strategies in Table 2.

Table 2 shows that under the baseline scenario, all the tested strategies are effective and cost-effective as opposed to no screening. A tested screening strategy is labeled "effective" if it yields more cumulative QALYs than the baseline strategy; and "cost-effective" if it yields an ICER smaller than the $50 K/QALY threshold for societal-willingness to pay (Hirth, Chernew, Miller, Fendrick, & Weissert, 2000). These findings are consistent with the current literature. In terms of effectiveness, the strategies are ranked in ascending order as one-time, routine 20, routine 10, conservative, 2nd current, 1st current, and aggressive. These results indicate that more intensive tests can derive more benefits. The cumulative cost values follow the same order. None of the strategies is dominated. In Table 3, we present the pairwise ICERs for all of the screening strategies evaluated. For example, the value 2597 in the left-hand corner means that it requires $2597 to gain one more QALY with the routine 20 strategy relative to the one-time strategy.

Table 2
Baseline result.

Policy	Cumulative QALYs	Cumulative costs	ICER vs. no screening
No screening	19.8642	1626	–
One-time	20.1983	1696	210
Routine 20	20.2570	1803	450
Routine 10	20.3329	2033	868
Conservative	20.3367	2071	941
2Nd current	20.3374	2085	970
1St current	20.3613	2302	1360
Aggressive	20.3700	2440	1610

Investigation of initial screening age, screening end age, and partial compliance

In this section, we test the impact of three important model parameters on the cost and effectiveness among the tested screening strategies. It should be noted that among the three parameters, the screening end age is least investigated in the literature. Figs. 3–5 compile the investigation results.

Fig. 3 shows that the effect of different initial screening ages (45, 50, and 55) on the tested strategies in terms of cumulative QALYs, cumulative costs, and ICERs versus no screening. Fig. 3a shows that for any initial screening age between 45 and 55 all the strategies result in more QALYs than no screening. In addition, for all the tested strategies, the cumulative QALYs decrease monotonically as we delay the initial screening. This decrease can be explained by the fact that the earlier the initial screening is performed, the more likely advanced adenomas are detected at early stages and thus the following polypectomy can reduce the CRC risk. Finally, comparing the four observation-based strategies (strategies 4–7 in the figure) demonstrates the effect on cumulative QALYs when varying the screening frequency for observation groups 2 and 3. Our results suggest that a strategy with more frequent colonoscopy testing for observation group 2 is more effective, which is reasonable as most people at approximately age 50 are found to have 1–2 small adenomas. On the other hand, the columns associated with the 2nd current and conservative strategies are nearly identical, which suggests that cumulative QALYs are insensitive to screening frequency changes for observation group 3. Fig. 3b shows that higher screening frequency results in an increase in the screening cost but a reduction in the treatment cost due to reduced CRC risk. Overall, the total cost will increase as the screening frequency increases. When comparing the four observation-based screening strategies, the results indicate that increased screening frequency for observation group 2 incurs a significant increase to the total cost, while varying screening interval for observation group 3 does not have much effect on the total cost. Fig. 3c shows that varying initial screening age does not change the ordering of the tested strategies in terms of the cost-effectiveness.

Fig. 4 presents the effect of different screening end ages (75, 80, and 85) on the tested strategies in terms of cumulative QALYs, cumulative costs, and ICERs versus no screening. For these analyses, we fixed the initial screening age to be 50. As Fig. 4 shows, varying screening end age does not have as significant an influence on the three measures as varying initial screening age does. Generally, as the termination of the screening is delayed, the QALYs tend to increase but the costs increase at the same time. The results do not change the ordering of the tested strategies in terms of the cost-effectiveness.

Table 3
Baseline incremental cost-effectiveness ratios (reported value compares policy reported in the column with the policy reported in the row).

Policy	ICERs ($/QALY)					
	Routine 20	Routine 10	Conservative	2Nd current	1St current	Aggressive
One-time	2597	1696	3866	2843	2782	4352
Routine 20	–	3288	9714	9200	8750	10,639
Routine 10	–	–	5069	3667	3571	5716
Conservative	–	–	–	9826	9875	13,875
2Nd current	–	–	–	–	11,000	10,871
1St current	–	–	–	–	–	10,875
Aggressive	–	–	–	–	–	–

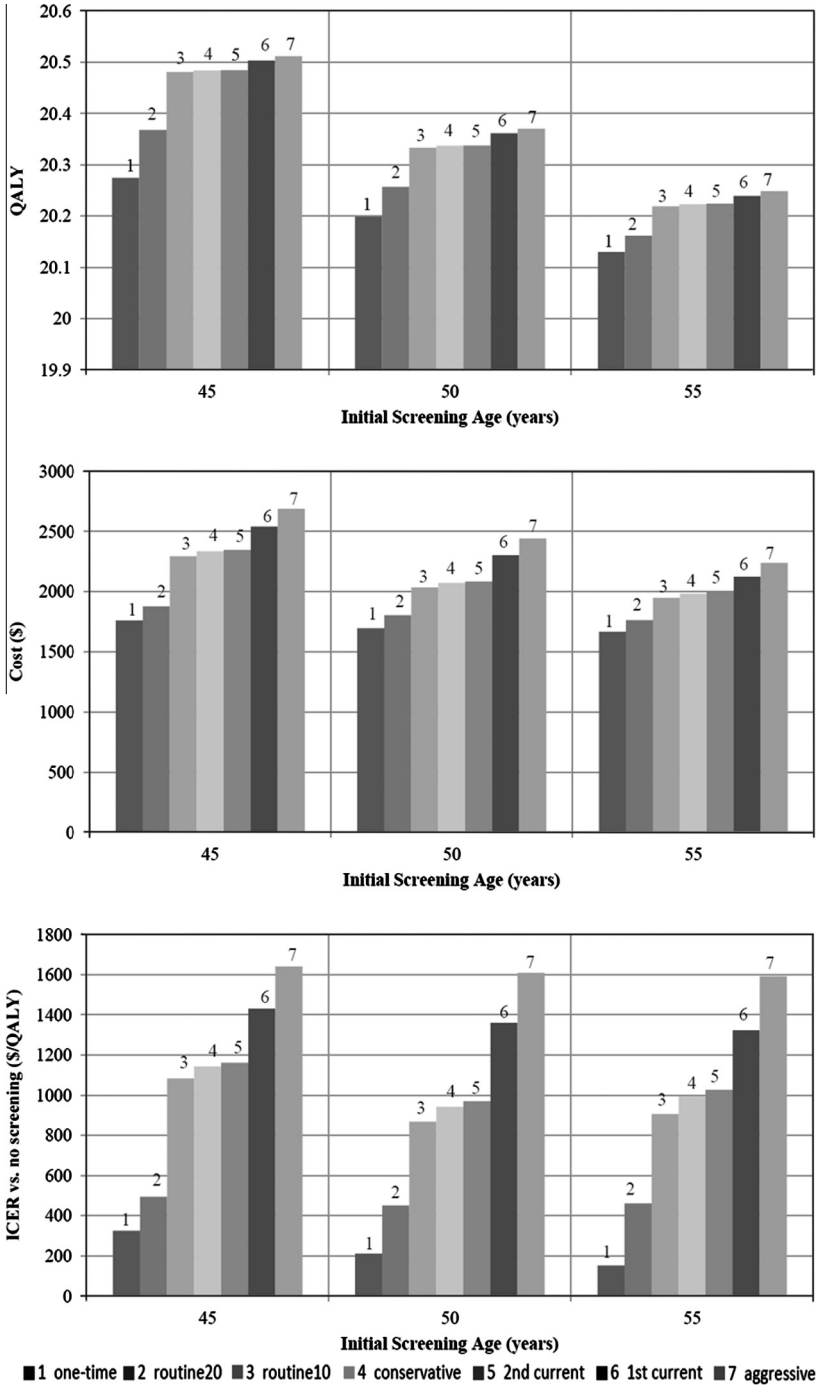

Fig. 3. Effect of different initial screening ages on the QALY, cost, and ICER vs. no screening.

Fig. 5 presents the effect of different compliance rates on the tested strategies. Both cumulative QALYs and cumulative costs would decrease as the compliance rate decreases, which is due to the extended screening intervals caused by partial compliance. In addition, for any tested screening strategy, the difference in each of the three measures would diminish as the compliance rate decreases, which implies that the effect of varying screening intervals is less significant with the decrease of the compliance rate.

Sensitivity analysis

To verify the robustness of our results, we performed both one-way and probabilistic sensitivity analyses on the cost and quality of life parameters. We acquired the ranges of plausible values for the analyzed parameters from Tafazzoli et al. (2009), as listed in Table 1.

In the first one-way sensitivity analysis, we grouped the cost of performing a diagnostic colonoscopy and the cost of performing a diagnostic colonoscopy followed by polypectomy for adenoma removal, since these two costs are highly positively correlated. We conducted a one-way sensitivity analysis by setting the two screening costs to their respective minimum and maximum values. In the second one-way sensitivity analysis, we grouped the costs for initial treatment of local CRC, regional CRC, and distant (metastasized) CRC, the yearly costs for continuing treatment of local CRC, regional CRC, and distant CRC, and the cost for terminal treatment

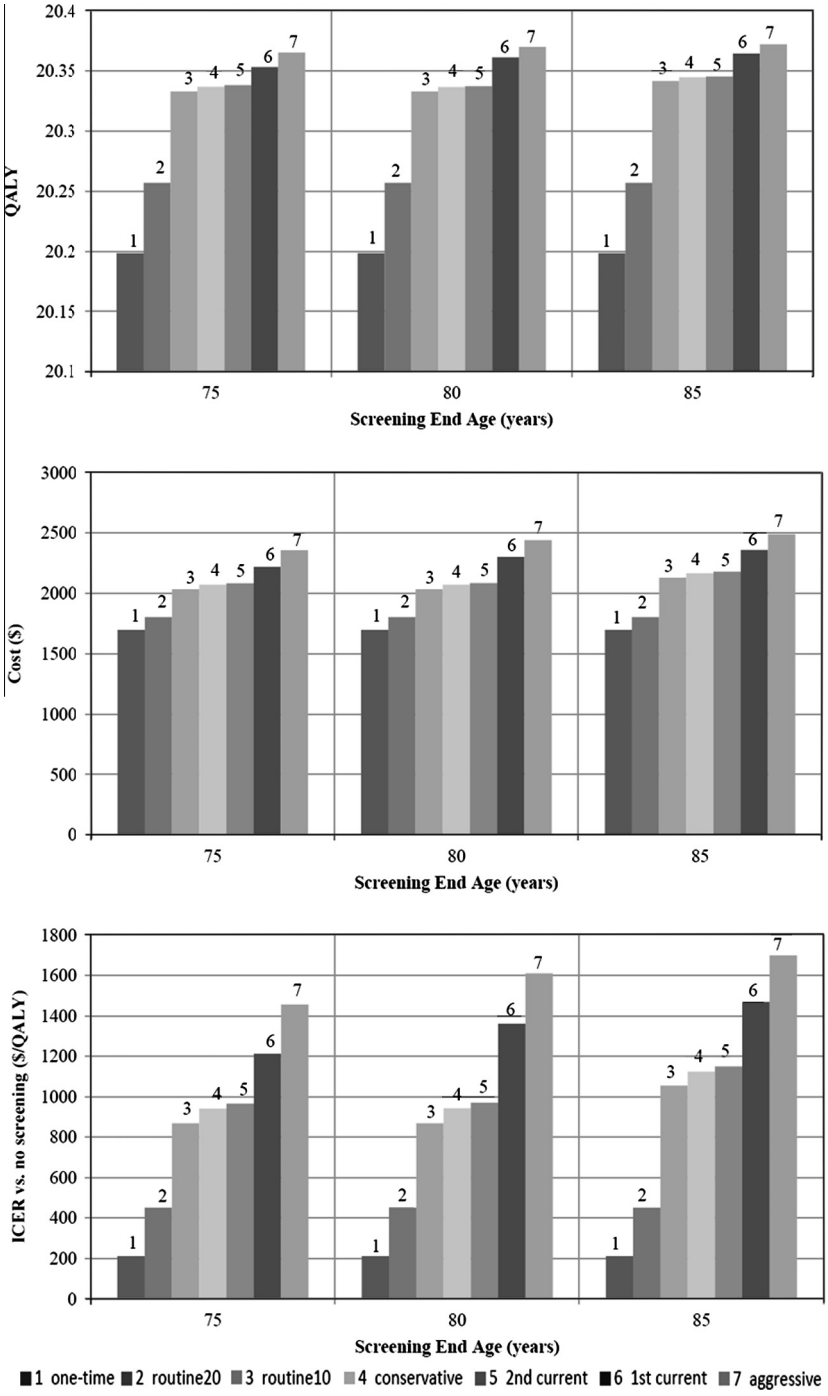

Fig. 4. Effect of different screening end ages on the QALY, cost, and ICER vs. no screening.

of CRC. We conducted the one-way sensitivity analysis for treatment costs in the same manner as for the screening costs. In Table 4, we report two ranges at each intersection of the upper triangle portion of the table. The first and second ranges in each intersection report the ranges of the ICERs for the one-way sensitivity analysis on the screening cost and treatment cost groups, respectively. For example, the two ranges (1.1–3.8) and (0.8–2.1) in the left-hand corner means that it requires $1100–$3800 to gain one more QALY with the routine 20 strategy relative to the one-time strategy if we vary screening costs; and it requires $800–$2100 to gain one more QALY for the same comparison if we vary treatment costs. From the table, we concluded that the comparative results among the tested screening strategies would be more sensitive to the screening costs as opposed to the treatment costs.

Similarly, we conducted three additional sensitivity analyses on the quality of life multipliers. We set the quality of life multipliers for adenoma states, local cancer, and regional cancer, to be their respective minimum and maximum values. We do not vary the value of the quality of life multiplier for distant cancer since such a range is not available in the literature. We report the analysis results in Table 5. From the table, we concluded that

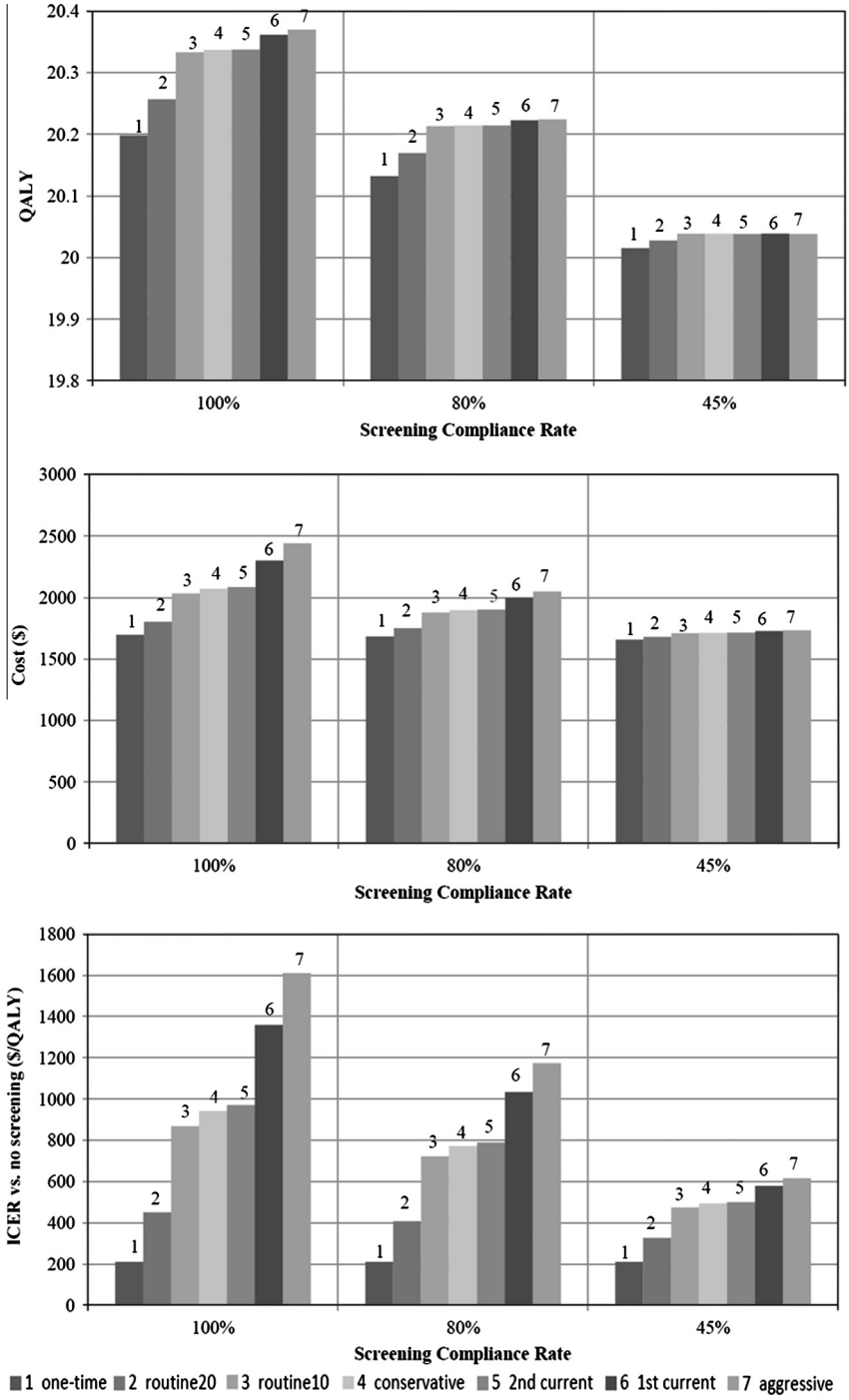

Fig. 5. Effect of different compliance rates on the QALY, cost, and ICER vs. no screening.

the comparative results among the tested screening strategies would be more sensitive to the quality of life multiplier for adenoma states than for other states. The fact that the comparative results are insensitive to the quality of life multipliers for various cancer states implies that the cancer location information may not need to be incorporated when assessing the screening strategies. In several cases, as the quality of life multiplier for adenoma states decreases to a certain level, some strategy would become dominated. For example, comparing the routine 10 and 2nd current strategies, when the quality of life multiplier for adenoma states decreases to 0.91, the routine 10 strategy would become dominated as fewer QALYs are derived and more costs are incurred from it than the 2nd current strategy. As we mentioned in section 'Baseline analysis', when applying the \$50 K/QALY threshold for societal-willingness to pay (Hirth et al., 2000), a strategy is not cost-effective if it yields an ICER greater than \$50 K/QALY with regard to the strategy it is comparing with. Thus, we observed from the right-hand bottom part of Table 5 that the 1st current and aggressive strategies may no longer be cost-effective compared to the routine 10, conservative, and

Table 4
Incremental cost-effectiveness ratio for the one-way sensitivity analysis on the cost parameters.

Policy	ICERs ($\times 10^3$ \$/QALY)					
	Routine 20	Routine 10	Conservative	2Nd current	1St current	Aggressive
One-time	(1.1–3.8) (0.8–2.1)	(2.1–5.6) (2.2–3.3)	(2.3–5.9) (2.4–3.5)	(2.4–6.0) (2.5–3.5)	(2.9–7.1) (3.3–4.2)	(3.4–8.0) (3.8–4.8)
Routine 20	–	(2.9–6.8) (3.2–4.1)	(3.1–7.3) (3.5–4.4)	(3.2–7.5) (3.6–4.5)	(3.9–8.9) (4.5–5.4)	(4.5–10.1) (5.3–6.1)
Routine 10	–	–	(9.2–19.3) (11.2–11.9)	(10.2–21.3) (12.5–13.1)	(7.9–16.6) (9.6–10.3)	(9.2–19.2) (11.3–11.9)
Conservative	–	–	–	(14.6–30.1) (17.9–18.7)	(7.7–16.2) (9.3–10.0)	(9.2–19.2) (11.3–11.9)
2Nd current	–	–	–	–	(7.4–15.7) (9.0–9.7)	(9.1–18.9) (11.1–11.7)
1St current	–	–	–	–	–	(13.2–26.8) (16.3–16.7)
Aggressive	–	–	–	–	–	–

Table 5
Incremental cost-effectiveness ratio for the one-way sensitivity analysis on the quality of life parameters.

Policy	ICERs ($\times 10^3$ \$/QALY)					
	Routine 20	Routine 10	Conservative	2nd current	1st current	Aggressive
One-time	(1.6–2.2) (1.8–1.9) (1.8–1.9)	(2.4–4.1) (2.9–3.1) (3.0–3.1)	(2.5–4.4) (3.1–3.3) (3.2–3.3)	(2.6–4.5) (3.2–3.4) (3.2–3.3)	(3.0–6.1) (3.9–4.1) (3.9–4.0)	(3.3–7.4) (4.4–4.7) (4.5–4.6)
Routine 20	–	(2.9–5.7) (3.8–4.0) (3.8–3.9)	(3.1–6.3) (4.0–4.3) (4.1–4.2)	(3.1–6.6) (4.1–4.4) (4.2–4.3)	(3.6–9.3) (5.1–5.3) (5.1–5.2)	(4.0–11.8) (5.8–6.1) (5.9–6.0)
Routine 10	–	–	(6.0–363.9) (11.7–12.1) (11.7–11.7)	(6.3-Dominated) (13.0–13.3) (13.0–13.0)	(5.4–79.7) (9.8–10.3) (10.0–10.1)	(5.9–448.8) (11.5–12.0) (11.7–11.8)
Conservative	–	–	–	(7.2-Dominated) (18.5–18.5) (18.5–18.5)	(5.3–69.6) (9.6–10.0) (9.7–9.9)	(5.9–460.9) (11.5–12.0) (11.6–11.7)
2nd current	–	–	–	–	(5.2–56.8) (9.3–9.7) (9.4–9.5)	(5.9–281.6) (11.3–11.8) (11.5–11.6)
1st current	–	–	–	–	–	(7.3-Dominated) (16.6–17.0) (16.6–16.6)
Aggressive	–	–	–	–	–	–

2nd current strategies. This implies that from a cost-effectiveness viewpoint, less aggressive and less frequent screening is more desirable.

The above one-way sensitivity analyses could not fully explore the effects of uncertainty. Hence, we further conducted a probabilistic sensitivity analysis on all the aforementioned model parameters. In our probabilistic sensitivity analysis, we applied the NORTA method to generate the model parameters (Cario & Nelson, 1998), which were treated as correlated random variables. Each random variable was generated by taking samples from a Beta distribution, which was generated based on the minimum and maximum values of the model parameter. In addition, we assumed that the standard deviation of the Beta distribution is one sixth of the plausible range of the model parameter. We further assumed that the model parameters within each group are of perfect positive correlation but the model parameters from different groups are of no correlation. In each replication of the probabilistic sensitivity analysis, we randomly sampled from the distributions, and evaluated each tested screening strategy based on the sampled model parameter values. We performed 200 replications for each tested strategy. We report the results in Fig. 6. Tafazzoli et al. (2009) conducted a similar probabilistic sensitivity analysis to compare a different set of CRC screening strategies.

In Fig. 6, each label represents one replication of a screening strategy. We observed that the routine 10, 2nd current, and conservative strategies largely overlap and thus do not present much difference in terms of the cost-effectiveness. Similarly, we observed that the 1st current and aggressive strategies do not differ much in terms of the cost-effectiveness. These probabilistic sensitivity analysis results are consistent with the results obtained from those one-way sensitivity analyses.

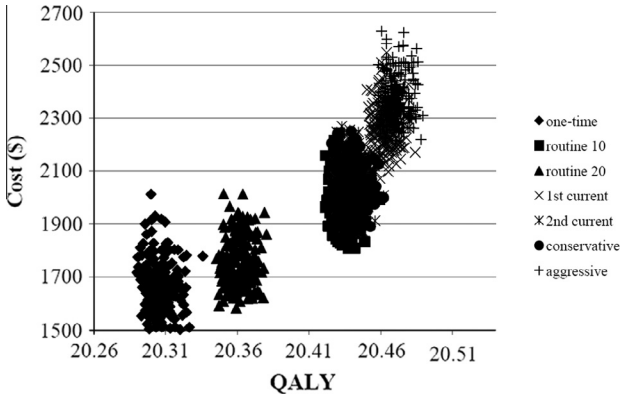

Fig. 6. Results for 200 replications of screening policies presented on the cost-effectiveness plane.

Conclusions and future work

In this paper, we develop a POMC model with detailed precancerous adenoma state description and demonstrate its

applicability in evaluating the cost-effectiveness of various colonoscopy screening strategies. Our natural history model is calibrated and validated using a large longitudinal clinical data set from a specific population cohort. The intelligent design of the belief state update procedure enables our model to accommodate partial observation of patients' health states and inaccuracy of colonoscopy screening and adenoma removal, which differentiates our model from most traditional Markov models. Through a comprehensive literature review of CRC screening models, we clearly identify our main contributions to the current literature. Methodologically, our model is the first POMC model analyzing cost-effectiveness of CRC screening strategies; and practically, we evaluate several observation-based colonoscopy screening strategies, which provides the possibility of improving the current colonoscopy screening guideline when general population data is available.

Our numerical studies show that the current screening guideline and its variations are cost-effective compared to the no screening strategy for our studied population cohort. In addition, we observe that varying the screening interval is influential on the cost-effectiveness, especially when varying the interval for observation group 2 (1–2 non-advanced adenomas). Our experimental results also suggest that using a larger interval for group 2 (e.g., 10 years) is likely to be more cost-effective than the current recommendation. Finally, our study indicates that special attention should be given to the estimation of the screening compliance rate and screening cost when assessing the cost-effectiveness of colonoscopy screening strategies.

There are three limitations in this paper. First, our numerical results and recommendations are not generalizable because we estimated our parameters based on a specific population cohort. However, the qualitative conclusions drawn from our model are consistent with those from other models. Second, our model does not incorporate other CRC risk factors such as gender, race, and CRC family history, which are likely to affect the CRC disease dynamics. Finally, we cannot efficiently optimize the strategy design, which requires the application of stochastic optimization, e.g., approximate dynamic programming. Our objective in this paper is to provide a POMC-based framework for more detailed modeling of colonoscopy screening interventions and thus open up many possibilities to the research community in applying POMC models to assess colonoscopy screening strategies with or without combination of other screening methods. A minor limitation is that it is impossible to determine the correlation coefficients within each group of model parameters and between different groups in the probabilistic sensitivity analysis. This limitation may decrease the value of the probabilistic sensitivity analysis.

In the future, we plan to address the above limitations by calibrating our POMC model with clinical data from additional and larger cohorts and incorporating additional CRC risk factors into our model. In addition, we plan to evaluate the cost-effectiveness of CRC screening methods other than colonoscopy. We also plan to develop a stochastic optimization method to efficiently select promising CRC screening strategies. Finally, since correlations are unknown, one should judiciously apply the conclusions drawn from our probabilistic sensitivity analysis.

Acknowledgements

We thank Dr. Eric Sherer from Indiana University Medical School for providing the CRC natural history model parameters and for his comments and edits of the manuscript. We also thank Dr. Jingyu Zhang from Philips Research North America for his insightful suggestions.

Appendix A

Definitions of precancerous adenoma states.

PCA No.	n_s	n_m	n_l	PCA No.	n_s	n_m	n_l	PCA No.	n_s	n_m	n_l
1	1	0	0	29	1	0	1	57	3	1	2
2	2	0	0	30	2	0	1	58	0	2	2
3	3	0	0	31	3	0	1	59	1	2	2
4	4	0	0	32	4	0	1	60	2	2	2
5	5	0	0	33	5	0	1	61	0	3	2
6	6	0	0	34	0	1	1	62	1	3	2
7	0	1	0	35	1	1	1	63	0	4	2
8	1	1	0	36	2	1	1	64	0	0	3
9	2	1	0	37	3	1	1	65	1	0	3
10	3	1	0	38	4	1	1	66	2	0	3
11	4	1	0	39	0	2	1	67	3	0	3
12	5	1	0	40	1	2	1	68	0	1	3
13	0	2	0	41	2	2	1	69	1	1	3
14	1	2	0	42	3	2	1	70	2	1	3
15	2	2	0	43	0	3	1	71	0	2	3
16	3	2	0	44	1	3	1	72	1	2	3
17	4	0	0	45	2	3	1	73	0	3	3
18	0	3	0	46	0	4	1	74	0	0	4
19	1	3	0	47	1	4	1	75	1	0	4
20	2	3	0	48	0	5	1	76	2	0	4
21	3	3	0	49	0	0	2	77	0	1	4
22	0	4	0	50	1	0	2	78	1	1	4
23	1	4	0	51	2	0	2	79	0	2	4
24	2	4	0	52	3	0	2	80	0	0	5
25	0	5	0	53	4	0	2	81	1	0	5
26	1	5	0	54	0	1	2	82	0	1	5
27	0	6	0	55	1	1	2	83	0	0	6
28	0	0	1	56	2	1	2				

PCA no.: precancerous adenoma state number; n_s: number of small adenomas; n_m: number of medium adenomas; and n_l: number of large adenomas.

Appendix B

CRC incidence rate and cause specific mortality rate for general population and studied cohort population.

Age	CRC incidence rate (per 100,000 individuals)		CRC cause specific mortality rate (per 100,000 individuals)	
	General population	Cohort population	General population	Cohort population
40	16	1777	4	811
45	29	2414	8	2059
50	55	3144	13	2764
55	76	3978	22	3543
60	108	4929	34	4433
65	165	5995	51	5438
70	217	7147	74	6545
75	272	8306	102	7700
80	327	9330	144	8790
85	358	10,068	217	9671

References

Arias, E. (2010). *United States Life Tables, national vital statistics reports* (Vol. 58(21)). Hyattsville, MD: National Center for Health Statistics. <http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58_21.pdf>.

Ayer, T., Alagoz, O., & Stout, N. K. (2012). A POMDP approach to personalize mammography screening decisions. *Operations Research, 60*(5), 1019–1034.

Cario, M., & Nelson, B. (1998). Numerical methods for fitting and simulating, autoregressive-to-anything processes. *INFORMS Journal on Computing, 10*(1), 72–81.

Cubbage, D. (2004). *Simulation of colorectal cancer: the natural history of disease*. Master's thesis, North Carolina State University

Frazier, A. L., Colditz, G. A., Fuchs, C. S., & Kuntz, K. M. (2000). Cost-effectiveness of screening for colorectal cancer in the general population. *Journal of the American Medical Association, 284*(15), 1954–1961.

Gold, M. R., Siegel, J., Russell, L., & Weinstein, M. (Eds.). (1996). *Cost-effectiveness in health and medicine*. NY, USA: Oxford University Press.

Hedden, L., Kennecke, H., Villa, D., Johnston, K., Speers, C., Kovacic, L., et al. (2012). Incremental cost-effectiveness of the pre-and post-bevacizumab eras of metastatic colorectal cancer therapy in British Columbia, Canada. *European Journal of Cancer, 48*(13), 1969–1976.

Heitman, S., Hilsden, R., Au, F., Dowden, S., & Manns, B. (2010). Colorectal cancer screening for average-risk North Americans: An economic evaluation. *PLoS Medicine, 7*(11), 1–13.

Hirth, R. A., Chernew, M. E., Miller, E., Fendrick, A. M., & Weissert, W. G. (2000). Willingness to pay for a quality-adjusted life year in search of a standard. *Medical Decision Making, 20*(3), 332–342.

Howlader, N., Noone, A. M., Krapcho, M., Neyman, N., Aminou, R., Altekruse, S. F., et al. (2012). *SEER cancer statistics review, 1975–2009 (vintage 2009 populations)*. Bethesda, MD: National Cancer Institute. <http://seer.cancer.gov/csr/1975_2009_pops09/> (based on November 2011 SEER data submission, posted to the SEER web site, April 2012)..

Jemal, A., Bray, F., & Center, M. (2011). Global cancer statistics, 2011. *CA Cancer Journal for Clinicians, 61*(2), 69–90.

Ladabaum, U., Song, K., & Fendrick, A. M. (2004). Colorectal neoplasia screening with virtual colonoscopy: When, at what cost, and with what national impact. *Clinical Gastroenterology and Hepatology, 2*(7), 554–563.

Lansdorp-Vogelaar, I., Ballegooijen, M., Zauber, A. G., Boer, R., Wilschut, J., & Habbema, J. D. F. (2009). At what cost will screening with CT colonography be competitive? A cost-effectiveness approach. *International Journal of Cancer, 124*(5), 1161–1168.

Levin, B., Lieberman, D. A., McFarland, B., Andrews, K. S., Brooks, D., Bond, J., et al. (2008). American Cancer Society Colorectal Cancer Advisory Group, U.S. Multi-Society Task Force, American College of Radiology Colon Cancer Committee, Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: A joint guideline from the American Cancer Society, the U.S. Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology, 134*(5), 1570–1595.

Loeve, F., Boer, R., van Oortmarssen, G. J., van Ballegooijen, M., & Habbema, J. D. (1999). The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. *Computers and Biomedical Research, 32*(1), 13–33.

Loeve, F., Brown, M. L., Boer, R., van Ballegooijen, M., van Oortmarssen, G. J., & Habbema, J. D. F. (2000). Endoscopic colorectal cancer screening: A cost-saving analysis. *Journal of the National Cancer Institute, 92*(7), 557–563.

Lucidarme, O., Cadi, M., Berger, G., Taieb, J., Poynard, T., Grenier, P., et al. (2012). Cost-effectiveness modeling of colorectal cancer: Computed tomography colonography vs colonoscopy or fecal occult blood tests. *European Journal of Radiology, 81*(7), 1413–1419.

Maillart, L. M., Ivy, J. S., Ransom, S., & Diehl, K. (2008). Assessing dynamic breast cancer screening policies. *Operations Research, 56*(6), 1411–1427.

National Cancer Institute (2012). Colorectal Cancer Model Profiles: Cancer Intervention and Surveillance Modeling Network (CISNET). <http://cisnet.cancer.gov/colorectal/profiles.html>.

Ness, R. M., Holmes, A. M., Klein, R., & Dittus, R. S. (2000). Cost-utility of one-time colonoscopic screening for colorectal cancer at various ages. *American Journal of Gastroenterology, 95*(7), 1800–1811.

Pignone, M., Russell, L., & Wagner, J. (Eds.). (2005). *Economic models of colorectal cancer screening in average-risk adults: Workshop summary*. Washington, DC: National Academies Press.

Rex, D. K., Cutler, C. S., Lemmel, G. T., Rahmani, E. Y., Clark, D. W., Helper, D. J., et al. (1997). Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology, 112*(1), 24–28.

Rex, D. K., Johnson, D. A., Anderson, J. C., Schoenfeld, P. S., Burke, C. A., & Inadomi, J. M. (2009). American College of Gastroentrology guidelines for colorectal cancer screening 2009. *American Journal of Gastroenterology, 104*(6), 1613.

Roberts, S. D., Wang, L., Klein, R., Ness, R. M., & Dittus, R. S. (2007). Development of a simulation model of colorectal cancer. *ACM Transactions on Modeling and Computer Simulation, 18*(1), 1–30.

Rutter, C. M., & Savarino, J. E. (2010). An evidence-based microsimulation model for colorectal cancer: Validation and application, cancer epidemiology. *Biomarkers & Prevention, 19*(8), 1992–2002.

Rutter, C. M., Zaslavsky, A. M., & Feuer, E. J. (2010). Dynamic microsimulation models for health outcomes: A review. *Medical Decision Making, 31*(1), 10–18.

Sherer, E. A., Imperiale, T. F., Ambedkar, S., Perng, S., Yih, Y. (2010). An adaptive–predictive model of colonic neoplasia at colonoscopy. In *Proceedings of the 4th international conference on population balance modeling, 15–17 September 2010*, Berlin, Germany

Sherer, E. A., Imler, T. D., & Imperiale, T. F. (2012). The effect of colonoscopy preparation quality on adenoma detection rates. *Gastrointestinal Endoscopy, 75*(3), 545–553.

Sobhani, I., Alzahouri, K., Ghout, I., Charles, D. J., & Durand-Zaleski, I. (2011). Cost-effectiveness of mass screening for colorectal cancer: Choice of fecal occult blood test and screening strategy. *Diseases of the Colon and Rectum, 54*(7), 876–886.

Song, K., Fendrick, A., & Ladabaum, U. (2004). Fecal DNA testing compared with conventional colorectal cancer screening methods: A decision analysis. *Gastroenterology, 126*(5), 1270–1279.

Sonnenberg, A., Delco, F., & Inadomi, J. M. (2000). Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Annals of Internal Medicine, 133*(8), 573–584.

Tafazzoli, A., Roberts, S., Ness, R., Klein, R., & Dittus, R. (2009). Probabilistic cost-effectiveness comparison of screening strategies for colorectal cancer. *ACM Transactions on Modeling and Computer Simulation, 19*(2), 1–29.

U.S. Preventive Services Task Force (2008). Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of Internal Medicine, 149*(9), 627–637.

van Rossum, L. G. M., van Rijn, A. F., Verbeek, A. L. M., van Oijen, M. G. H., Laheij, R. J. F., Fockens, P. F., et al. (2011). Colorectal cancer screening comparing no screening, immunochemical and guaiac fecal occult blood tests: A cost-effectiveness analysis. *International Journal of Cancer, 128*(8), 1908–1917.

Vijan, S., Hwang, E. W., Hofer, T. P., & Hayward, R. A. (2001). Which colon cancer screening test? A comparison of costs, effectiveness, and compliance. *American Journal of Medicine, 111*(8), 593–601.

Wilschut, J. A., Hol, L., Dekker, E., Jansen, J. B., van Leerdam, M. E., Lansdorp-Vogelaar, I., et al. (2011). Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. *Gastroenterology, 141*(5), 1648–1655.

Zauber, A. G., Winawer, S. J., O'Brien, M. J., Lansdorp-Vogelaar, I., Ballegooijen, M., Hankey, B. F., et al. (2012). Colonoscopic polypectomy and long-term prevention of colorectal cancer deaths. *New England Journal of Medicine, 366*(8), 687–696.

Zauber, A. G., Lansdorp-Vogelaar, I., Knudsen, A. B., Wilschut, J., van Ballegooijen, M., & Kuntz, K. M. (2008). Evaluating test strategies for colorectal cancer screening: A decision analysis for the U.S. Preventive Services Task Force. *Annals of Internal Medicine, 149*(9), 659–669.

Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., & Inman, B. A. (2012a). Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management, 14*(4), 529–547.

Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., & Inman, B. A. (2012b). Optimization of PSA screening policies a comparison of the patient and societal perspectives. *Medical Decision Making, 32*(2), 337–349.