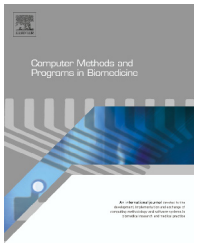
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 112 (2013) 563–579

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



A predictive model of longitudinal, patient-specific colonoscopy results

Eric A. Sherer[a,*], *Sanmit Ambedkar*[b], *Sally Perng*[b], *Yuehwern Yih*[b], *Thomas F. Imperiale*[a,c,d]

[a] *Roudebush Veterans Affairs Medical Center, Indianapolis, IN, USA*
[b] *Purdue University, West Lafayette, IN, USA*
[c] *Indiana University Medical Center, Indianapolis, IN, USA*
[d] *Regenstrief Institute, Indianapolis, IN, USA*

ARTICLE INFO

Article history:
Received 3 June 2011
Received in revised form
17 April 2013
Accepted 10 July 2013

Keywords:
Biomedical engineering
Colorectal cancer
Dynamic simulation
Mathematical modeling
Population balance
Patient specific

ABSTRACT

We suggest a model framework, in which an individual patient's risk for colonic neoplasia varies based on findings from his previous colonoscopies, to predict longitudinal colonoscopy results. The neoplasia natural history model describes progression through four neoplasia development states with patient age. Multiple natural history model parameter sets are assumed to act concurrently on the colon and parameter set prevalence combinations, whose a priori likelihoods are a function of patient sex, provide a basis set for patient-level predictions. The novelty in this approach is that after a colonoscopy, both the parameter set combination likelihoods and their model predictions can adjust in a Bayesian manner based on the results and conditions of the colonoscopy. The adjustment of model predictions operationalizes the clinical knowledge that multiple or advanced neoplasia at baseline colonoscopy is an independent predictor of multiple or advanced neoplasia at follow-up colonoscopy – and vice versa for negative colonoscopies – and the adjustment of parameter set combination likelihoods accounts for the possibility that patients may have different neoplasia development rates. A model that accurately captures serial colonoscopy results could potentially be used to design and evaluate post-colonoscopy treatment strategies based on the risk of individual patients. To support model identification, observational longitudinal colonoscopy results, procedure details, and patient characteristics were collected for 4084 patients. We found that at least two parameter sets specific to each sex with model adjustments was required to capture the longitudinal colonoscopy data and inclusion of multiple possible parameter set combinations, which account for random variations within the population, was necessary to accurately predict the second-time colonoscopy findings for patients with a history of advanced adenomas. Application of this model to predict CRC risks for patients adhering to guideline recommended follow-up colonoscopy intervals found that there are significant differences in risk with patient age, gender, and preparation quality and demonstrates the need for a more rigorous investigation into these recommendations.

Published by Elsevier Ireland Ltd.

* *Corresponding author*. Tel.: +1 318 257 3236.
E-mail address: ersherer@iupui.edu (E.A. Sherer).
0169-2607/$ – see front matter. Published by Elsevier Ireland Ltd.
http://dx.doi.org/10.1016/j.cmpb.2013.07.007

1. Introduction

The development of colorectal cancer (CRC) is somewhat unique among cancers in that precancerous neoplasia (i.e. adenomas), can be both identified and removed during colonoscopy. As such, colonoscopy not only detects cancer but the removal of adenomas short-circuits and prevents the development of potential colorectal cancer [1]. But cancer prevention via colonoscopy is not uniform across the patient population. Clinical studies of baseline and follow-up colonoscopy have found that the presence of multiple or advanced adenomas at the initial colonoscopy is independently related to an increased incidence of colon cancer and advanced adenoma at follow-up [2–6] while a negative initial colonoscopy is associated with lower risk of CRC [7]. Additionally, baseline adenoma prevalence has been shown to have a stronger connection to subsequent advanced neoplasia than external risk characteristics such as race, gender, and family and smoking history [8]. As such, current colonoscopy surveillance guidelines are stratified based on the adenoma findings at baseline, irrespective of procedure indication [9]. The objective of this study is to identify a mathematical framework and model that accurately predicts adenoma findings at first-time colonoscopy and then utilizes an individual patient's findings to make accurate predictions for follow-up colonoscopies. A mathematical model that accurately captures the dynamics adenoma formation and transformation to colorectal cancer could potentially be used to design or evaluate post-colonoscopy treatment based on the dynamics of post-colonoscopy risk for colorectal cancer on an individual which are known to vary with the result of a baseline colonoscopy.

Up to 95% of malignant colorectal cancers are thought to develop from small, benign tumors called adenomas [10]. Adenomas are an intermediate and benign cancer development state in which genetic changes produced neoplastic tissue but further changes are necessary for tumor initiation and progression [11–14]. Numerous mutations have been identified in adenomas and carcinomas with considerable variation in intra- as well as inter-neoplasia mutation characteristics [15–17]. In addition, growth conditions (e.g. vascularization, depth of neoplasia and surface conditions, or local chemistry) may influence the rate of expansion and progression of neoplasia. The macroscopic implications of the genetic and growth condition requirements and their heterogeneity are twofold. First, the development of neoplasia and their progression from adenoma to CRC requires time: CRC incidence increases with patient age and lags behind adenoma incidence [18]. And secondly, individual neoplasia can have variable rates of growth and progression.

Discrete-event simulation and micro-simulation models have examined the population-level effects of colonoscopy [19–22], but these models have several limitations when applied to predicting post-colonoscopy risks. The first limitation is that post-colonoscopy model predictions have not been validated; validation is essential for accurately capturing colonoscopy effects. In this study, models are fit to data for up to three serial colonoscopies to capture these behaviors. The second limitation is that, to varying degrees, previous models combine adenoma details together and focus of colorectal cancer incidence. However, such details are important when describing CRC development following a colonoscopy where subsequent risk for CRC is determined by the combination of growth of adenomas missed by the initial procedure and the development and progression of new adenomas. In this study, in addition to CRC, the models are tuned to and evaluated on capturing the number and size of adenomas found at colonoscopy. The final limitation is these models are population-level models and to not include the potential for inter-individual variability in neoplasia growth. In this manuscript, a model framework is introduced that accounts for the possibility of differences between patients and adjusts the model predictions of an individual patient in a Bayesian manner based previous colonoscopy results. This Bayesian approach requires the complete probability distribution of serial colonoscopy results for each patient but, because of the number of possible colonoscopy results and the serial colonoscopies, the number of trials to approximate the probabilities for each patient using multiple trials with a discrete-event or micro-simulation becomes infeasible. Therefore, a computational method is described for calculating the complete probability distribution of serial colonoscopy results.

Colonoscopy is the only colon measurement mode which visualizes the entire length of the colon, detects adenomas less than 10 mm, and removes polyps for pathological analysis. In addition, colonoscopy is the gold standard for neoplasia detection and has a high sensitivity for neoplasia [23]. For these reasons, we rely on colonoscopy data to provide guidance for estimates of the neoplasia truly present in patients.

2. Methods

2.1. *Model framework*

An overview of the model framework is as follows (see Fig. 1). The adenoma natural history progression model is the basic unit of the mathematical framework (Section 2.1.1) where application of a parameter set to this model governs formation and progression of an individual adenoma toward CRC. Patient-level predictions are a combination of predictions from multiple natural history model parameter sets which act simultaneously and independently. A series of parameter set combinations provides a basis set of patient-level predictions where individual patient's demographic characteristics determine the a priori likelihood of each parameter set combination in the basis set (Section 2.1.2) where the model predictions can adjust in a Bayesian manner based on the colonoscopy findings (Section 2.1.5). The rationale behind this adaption is to account for the possibility that patients may be different (i.e. subject to different rate functions) and variation in the population may also be due to differences between patients.

2.1.1. *Natural history model*

Similar to the MISCAN-COLON model [24], the colorectal adenoma to carcinoma sequence was categorized into four states: diminutive (≤5 mm) adenoma, medium (6–9 mm) adenoma, large/advanced (≥10 mm or with villous or high-grade dysplasia histology) adenoma, and cancer (see Fig. 1). We assumed that the rate of appearance of new diminutive adenomas