Article:
Rose, J, Homa, L, Kong, CY et al. (9 more authors) (2019) Development and Validation of a Model to Predict Outcomes of Colon Cancer Surveillance. Cancer Causes and Control, 30 (7). pp. 767-778. ISSN 0957-5243

https://doi.org/10.1007/s10552-019-01187-x

Development and Validation of a Model to Predict Outcomes of Colon Cancer Surveillance

Johnie Rose[1,2], Laura Homa[1], Chung Yin Kong[3], Gregory S. Cooper[1,2,4], Michael W. Kattan[2,5], Bridget Ermlich[6], Jeffrey P. Meyers[7], John N. Primrose[8], Sian A. Pugh[8], Bethany Shinkins[9], Uriel Kim[1], Neal J. Meropol[1,2,10]

1. Case Western Reserve University School of Medicine, Cleveland, OH, USA; 2. Case Comprehensive Cancer Center, Cleveland, OH, USA; 3. Massachusetts General Hospital Institute for Technology Assessment, Cambridge, MA, USA; 4. University Hospitals Seidman Cancer Center, Cleveland, OH, USA; 5. Cleveland Clinic Foundation, Cleveland, OH, USA; 6. University Hospitals Cleveland Medical Center, Cleveland, OH, USA; 7. Mayo Clinic Minnesota, Rochester, MN, USA; 8. University Surgery, Cancer Sciences, University of Southampton, Southampton, UK; 9. Test Evaluation Group, University of Oxford, Oxfordshire, UK; 10. current affiliation Flatiron Health, New York, NY

Word count: 3,944; Tables: 2; Figures: 5; Appendices: 3

A poster describing a preliminary version of this work was presented at the Society for Medical Decision Making Meeting Annual Meeting in St. Louis, MO, USA on October 21, 2015.

Financial support for this study was provided by a Mentored Research Scholar Grant in Applied and Clinical Research, MRSG-13-315-01-CPHPS, from the American Cancer Society (JR); by a National Cancer Institute Program Grant, 5P30 CA043703-21 (JR, MWK, GSC, and NJM); and by the Clinical and Translational Science Collaborative of Cleveland, UL1TR000439 from the National Center for Advancing Translational Sciences (NCATS) component of the National Institutes of Health and NIH roadmap for Medical Research (JR). The Follow-up After Colorectal Surgery (FACS) trial was funded by the UK National Institute for Health Research Health Technology Assessment Programme (JNP, SAP, BS). The funding agreements ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the funders.

Acknowledgments

- Daniel J. Sargent, PhD – Mayo Clinic, Minnesota – For assistance with Clinical Outcomes of Surgical Therapy (COST) Trial data
- Alan P. Venook, MD – University of California San Francisco – For input regarding clinical practice norms
- Steven Cohen, MD – Abington Memorial Hospital/Jefferson Medical College Francisco – For input regarding clinical practice norms

Corresponding author: Johnie Rose

johnie.rose@case.edu

216-368-6860

ORCID ID: 0000-0001-7458-9303

Running title: Model to Predict Outcomes of Colon Cancer Surveillance

Key words: colon cancer, colorectal cancer, recurrence, surveillance, model

ABSTRACT

PURPOSE: Clinical trials suggest that intensive surveillance of colon cancer (CC) survivors to detect recurrence increases curative-intent treatment, though any survival benefit of surveillance as currently practiced appears modest. Realizing the potential of surveillance will require tools for identifying patients likely to benefit and for optimizing testing regimens. We describe and validate a model for predicting outcomes for any schedule of surveillance in CC survivors with specified age and cancer stage.

METHODS: A Markov process parameterized based on individual-level clinical trial data generates natural history events for simulated patients. A utilization submodel simulates surveillance and diagnostic testing. We validate the model against outcomes from the Follow-up After Colorectal Surgery (FACS) trial.

RESULTS: Pre-validation sensitivity analysis showed no parameter influencing curative-intent treatment by >5.0% or OS5 by >1.5%. In validation, the proportion of recurring subjects predicted to receive curative-intent treatment fell within FACS 95% confidence intervals for carcinoembryonic antigen (CEA)-intensive, computed tomography (CT)-intensive, and combined CEA+CT regimens, but not for a minimum surveillance regimen, where the model overestimated recurrence and curative treatment. Observed five-year overall survival (OS5) fell within 95% prediction intervals for all regimens.

CONCLUSION: The model performed well in predicting curative surgery for three of four FACS arms. It performed well in predicting OS5 for all arms.

INTRODUCTION

More than three quarters of the nearly 95,000 Americans diagnosed annually with colon cancer (CC) receive curative-intent treatment.[1] While the majority will be cured in the long term, approximately one fifth will experience recurrence, consisting of disease relapse or development of a new primary (metachronous) CC[2], with median survival of two to three years.[3][4]

Most CC survivors undergo post-treatment surveillance involving periodic follow-up testing to detect recurrence at an earlier stage where salvage surgery can provide a cure. Two previous meta-analyses showed improved mortality with intensive surveillance (i.e. more frequent testing using multiple modalities) compared to less intensive surveillance (risk ratio 0.73, 95% CI 0.60 to 0.89[5]; odds ratio 0.73, 95% CI 0.59 to 0.91[6]). However, a more recent meta-analysis[7], which included three large trials published in the last decade[2, 8, 9], showed no survival benefit from intensive surveillance. Across all published surveillance trials, the strategies tested, the populations studied, and the study periods vary significantly.[2, 9–15] Unsurprisingly, an evidence-based consensus among expert panels regarding how best to follow these patients is lacking.[16–22]

Part of the frustrating lack of progress in identifying an approach to surveillance that improves survival may arise from the heterogeneity of patient characteristics, disease features, and treatment histories. A number of factors have been found to impact recurrence risk in CC, including age, cancer stage, preoperative carcinoembryonic antigen (CEA) level, tumor differentiation, perineural and lymphovascular invasion, number of lymph nodes harvested at primary surgery, number of cancer-containing nodes, whether the patient received adjuvant chemotherapy, microsatellite instability, and others.[23–27] Notably, existing guidelines do not offer algorithms for tailoring surveillance based on a patient's specific risk factors or treatment history. While some guidelines incorporate a degree of flexibility in their recommendations based on whether patients are considered "high risk", they do not

offer objective criteria for defining high risk.[20, 22] This has led numerous authors to suggest a need for greater risk stratification and personalization of surveillance.[28–30]

For these reasons, there is an unmet need to develop a new approach for post-treatment surveillance—one which enables consideration of patient, disease, and treatment factors to tailor the surveillance regimen of each patient. Such multi-dimensional risk stratification may help identify the patients most likely to benefit from intensive surveillance and inform which surveillance strategies (specific tests and frequency) are most likely to result in detection of recurrence at a point where curative treatment is possible while minimizing unnecessary testing. The sheer number of distinct combinations and schedules of surveillance tests coupled with the number of risk-modifying patient, disease, and treatment factors make it unlikely that any number of traditional surveillance trials by themselves will ever provide the insight needed to personalize colon cancer surveillance.

We previously developed and validated a simple proof-of-concept simulation model based on published aggregate data that predicted the rate of curative salvage surgery and survival outcomes among colorectal cancer patients.[31] The model simulated the interplay between colorectal cancer recurrence natural history and early detection of recurrence through surveillance testing—an approach designed to allow the simulation of hypothetical surveillance strategies. Here, we build upon that work by describing the development and validation of the Colon Cancer Surveillance and Recurrence Model (CCSuRe). CCSuRe utilizes a more realistic representation of underlying disease processes (including non-constant progression risks over time and the effects of age and cancer stage on risk) that is fitted or "trained" based on individual-level clinical trial data. In addition, it uses a more complete model of surveillance and follow-up diagnostic testing. We validate the model by comparing observed outcomes from colon cancer patients in each arm of the recent Follow-up After Colorectal Surgery (FACS) trial[2] to CCSuRe-predicted outcomes for matched groups of simulated patients. This work represents the first independent validation of a colon cancer surveillance model against the results of a modern clinical trial.

It is a key step toward our objective of developing a surveillance model that can inform practice recommendations and personalize surveillance strategies for individual CC patients (and, eventually, for rectal cancer patients) based on patient and disease characteristics. Moving from one-size-fits-all surveillance toward a more tailored approach has the potential to increase the effectiveness and cost-effectiveness of surveillance while minimizing unnecessary testing.

METHODS

Overview of Approach

CCSuRe consists of two interacting submodels: a disease progression submodel and a utilization submodel. The former generates natural history events for simulated patients over time from the point of initial treatment of CC, and the latter simulates scheduled surveillance testing, follow-up diagnostic testing and recurrence treatment, which may affect survival. After clinical trial data is used to "train" the disease progression submodel, any hypothetical post-treatment surveillance regimen can be simulated by changing the parameters of the utilization submodel that control surveillance testing schedules.

The Disease Progression Submodel

We use a multi-state, non-homogenous Markov process to represent disease progression in continuous time.[32–36] There are four progressively more severe states: 1) no known recurrence, 2) detectable and resectable recurrence, 3) detectable recurrence which is not resectable, and 4) death (possible transitions depicted in **Figure A1** of **Appendix 1**). Transition intensities $\lambda_{ij}(t)$ describe the hazard rates (may be non-constant) at which transitions from state *i* to state *j* occur over time.

In practice, one can only infer that state transitions have occurred after the fact through clinical observation (with error). To infer parameters of the transition intensity functions, we used a maximum likelihood estimation process (**Appendix 2**) based on real-world follow-up data from the Clinical Outcomes of Surgical Therapy (COST) trial.[37] The COST trial enrolled 788 subjects across 48 U.S. sites with curatively treated Stage I-III colon cancer from 1994 to 1999. Participants underwent either laparoscopically-assisted or open surgical resection. With a median follow-up of 4.4 years, the COST investigators found no difference between arms in the primary endpoint of time to tumor recurrence.[37]

The distribution of recurrence sites was also based on the COST trial data using non-mutually exclusive categories of local extraluminal, intraluminal (metachronous), lung, liver, abdominal metastases, and other metastases. Patients with local extraluminal, intraluminal, lung, or liver recurrence are assumed to spend some amount of time in state 2 (detectable and resectable) before eventually advancing to state 3. However, patients whose recurrences manifest solely as abdominal metastases or "other" metastases transition from state 1 directly to state 3; such a patient would at no point be a candidate for curative treatment once their recurrent disease was detectable.

For "recurrence transitions" (from state 1 to 2 and from state 1 to 3), we assume that the transition intensities $\lambda_{12}(t)$ and $\lambda_{13}(t)$ are equal. The COST trial data demonstrated that the hazard of recurrence was non-constant over time, so we used a piecewise (first increasing, then decreasing) Weibull hazard. We also include the covariate of stage, assuming proportional hazards, because the hazard of recurrence differs based on cancer stage.

The transition from state 2 to 3 represents progression from a detectable and resectable recurrence to a non-resectable one. We model $\lambda_{23}(t)$ using a decreasing Weibull hazard since later recurrences tend to be more indolent.[38–41]

Transitions from states 1 and 2 to state 4, and from state 3 to 4, represent deaths. The two former transitions both represent death from other causes; we assume that $\lambda_{14}(t) = \lambda_{24}(t)$. The transition from State 3 to 4 represents death from cancer or other causes; $\lambda_{34}(t) > \lambda_{14}(t)$. Based on empirical hazard plots observed in the COST data, we assumed constant hazard and included an age covariate for all transitions into state 4.

After estimating the disease progression parameters, we estimated the rate of symptom onset among recurring patients. We assume a constant hazard for symptom onset; patients become eligible to develop symptoms when they transition from state 1 to states 2 or 3. To develop this hazard estimate, we used a chart review dataset of 62 recurring patients at our own institution (since COST trial

data does not describe symptomology at presentation), containing details of all relevant healthcare encounters following initial CC treatment and symptomology at each visit.

Survival of recurring patients after curative surgery

We fit an exponential survival model to the subset of COST patients who underwent curative surgery for recurrence using initial surgery as time 0 and death or last follow up as an event or point of censoring, respectively. This model was used to probabilistically assign life expectancies after curative treatment of recurrence.

Appendix 1 provides a detailed description of the disease progression submodel. Maximum likelihood calculations for parameter estimation were performed using MATLAB 2015a (Natick, MA).

The Utilization Submodel

Once the disease progression submodel was specified, it was coupled with a utilization submodel. The latter, a microsimulation developed using Anylogic v7.2 (Chicago, IL), re-creates scheduled surveillance testing, follow-up diagnostic testing, and treatment of recurrence for each of a series of simulated CC survivors for a surveillance period of up to five years. A schedule for surveillance carcinoembryonic antigen (CEA) testing, computed tomography (CT) of chest/abdomen/pelvis, and colonoscopy can be specified. The results of each test depend upon test sensitivity and specificity (**Table 1**), and upon the true state of simulated individuals in the disease progression submodel. Since test findings may be equivocal early in the course of a recurrence, causing a delay in diagnosis, we chose to model this phenomenon. The probabilities of equivocal results given the presence of disease (Table 1) were estimated based on the same dataset used to estimate symptom parameters and were varied extensively in sensitivity analysis.

Algorithms for repeat or follow-up diagnostic testing triggered by an equivocal or positive result were based on NCCN colon cancer guidelines[20], where possible, and on consensus from the coauthors and a panel of three medical oncologists with expertise in colorectal cancer prevention and treatment at separate academic institutions. **Appendix 3** depicts all testing algorithms.

Base Case and Sensitivity Analysis

We modeled a base case in which a population of 100,000 seventy-year-old CC survivors with a distribution of stage I, II, and III disease representative of U.S. patients (31.1%, 35.0%, and 33.9%, respectively)[42] underwent NCCN-adherent surveillance[20]. The regimen consisted of CEA measurement every three months for the first two years following initial surgery, then every six months for the following three years; annual CT of chest, abdomen, and pelvis for five years; and colonoscopy at years one and four. We conducted one-way sensitivity analysis on all utilization submodel parameters (Table 1) according to the ranges shown. We then simultaneously varied combinations of these parameters to which the model was sensitive. The primary outcomes of interest in sensitivity analyses were proportion of subjects undergoing curative salvage surgery for recurrence and overall survival at five years from the point of recurrence diagnosis (OS5). Recurrence proportion (the proportion of subjects experiencing recurrence) was not examined as an outcome in sensitivity analysis since it depends on disease progression parameters estimated by fitting to trial data as described above.

External Validation

After fitting the disease progression submodel to COST trial data and testing CCSuRe's sensitivity to variation in utilization parameters, we examined the model's ability to predict cancer-related outcomes for an independent group (i.e. a group whose data did not inform model development). We compared the actual experience of CC subjects in each of the four arms (and in all arms combined) of

the recent FACS trial to the outcomes predicted by CCSuRe for corresponding groups with the same sample size, mean age, and stage distribution. At 39 centers in the U.K. between 2003 and 2009, the FACS trial randomized 841 patients surgically treated for Dukes Stage A, B, or C colon cancer (corresponding to American Joint Committee on Cancer Stages I, II, and III[43]) to four arms undergoing either CEA based surveillance, CT based surveillance, combined CEA and CT, or minimum surveillance.[2] Median follow-up across all groups was 60 months; **Table 2** describes characteristics and assigned surveillance regimens for CC subjects in each FACS arm.

Outcomes examined in validation were recurrence proportion, proportion of recurring subjects undergoing curative salvage surgery, and OS5. Proportion of recurrences discovered because of symptoms was not examined given the small number of subjects in this category (36 CC subjects total).

For each of the four surveillance regimens, we ran CCSuRe 1,000 times using the number of subjects enrolled in the corresponding FACS arm (Table 2). For each outcome, we used the range of model outputs from the 2.5^{th} to 97.5^{th} percentile to define 95% prediction intervals.[44] We also constructed a calibration plot comparing observed to mean model-predicted survival of patients with recurrence at years one through five to assess model calibration and refinement.[45, 46]

The funding sources played no role in the design of the study or preparation of this manuscript.

RESULTS

Baseline Model Outputs

In the base case scenario modeling NCCN-adherent surveillance, the model predicted that 16.9% of survivors would experience recurrence and that 44.2% of recurring patients would be expected to undergo curative-intent treatment. OS5 for recurring patients was predicted to be 22.6%.

Sensitivity Analysis

Using the base case as a starting point, we conducted a sensitivity analysis in order to identify the utilization inputs to which model results were most sensitive. **Figure 1a** depicts the utilization parameters which, when varied individually from their baseline levels across the ranges shown in Table 1, most affected the proportion undergoing potentially curative treatment of recurrence. This outcome was most sensitive to positron emission tomography (PET) scan sensitivity. Varying PET sensitivity from 0.72 to 1.00 resulted in variation of the proportion treated curatively from 40.4% to 45.4%. **Figure 1b** depicts the four parameters which most affected predicted OS5. Again, PET sensitivity was most influential, with inputs between 0.72 and 1.00 resulting in OS5 ranging from 21.5% to 23.0%, respectively.

When CT scan sensitivity for all anatomical sites was simultaneously decreased to the lowest extremes, the model predicted a decrease in curative treatment proportion to 42.5%. At the highest extremes of CT sensitivity, curative treatment proportion rose to 44.3%. OS5 varied from 21.8% to 22.8%, respectively, for the low and high extremes of CT sensitivity. Since certain individual repeat/follow-up diagnostic testing wait times were influential, we ran the model with all such intervals simultaneously set to their low or high extremes. With all intervals at their lowest extremes (halved), curative treatment proportion rose to 46.9%, and OS5 to 23.6%. Increasing all intervals to their highest extremes (doubling) lowered curative treatment proportion and OS5 to 39.8% and 21.5%.

External Validation

The observed proportion recurring in each FACS trial arm, and in all arms combined, is shown in **Figure 2** alongside corresponding proportions predicted by the model. Predicted recurrence for the three intensive surveillance arms (CEA based, CT based, and CEA+CT based surveillance) fell within observed FACS 95% confidence intervals. Specifically, 15.2% (95% confidence interval [CI] 10.7% to 20.7%) of FACS subjects in the CEA arm were diagnosed with recurrence compared with a model-predicted 17.7% (95% prediction interval 12.9% to 23.0%); 20.3% (95% CI 15.0% to 26.5%) of FACS CT arm subjects were diagnosed with recurrence compared with a predicted 17.7% (95% prediction interval 12.4% to 23.3%); 13.9% (95% CI 9.5% to 19.4%) of FACS CEA+CT subjects were diagnosed with recurrence compared with a predicted 17.7% (95% prediction interval 12.5% to 22.6%). Diagnosed recurrence in the FACS minimum surveillance arm (10.8%; 95% CI 6.9% to 15.7%) was substantially lower than that in other arms and was lower than the model's prediction (17.0%; 95% prediction interval of 12.1% to 22.0%). The model's prediction for the recurrence proportion across all arms combined (17.4%; 95% prediction interval 15.1% to 20.1%) was slightly higher than observed in FACS (15.0%; 95% CI 12.6% to 17.6%), but still fell within the corresponding FACS 95% CI.

Figure 3 compares the observed proportion of patients diagnosed with recurrence who underwent curative-intent treatment with corresponding predicted levels. For the CEA, CT, and CEA+CT arms, CCSuRe predictions fell within FACS 95% CI's. Specifically, 33.3% (95% CI 18.2% to 52.6%) of FACS CEA arm subjects diagnosed with recurrence underwent curative-intent treatment compared with a model-predicted 41.9% (95% prediction interval 26.3% to 57.9%); 46.3% (95% CI 31.7% to 62.4%) of FACS CT arm subjects diagnosed with recurrence underwent curative-intent treatment compared with a model-predicted 40.1% (23.3% to 56.8%); 37.9% (95% CI 20.7% to 56.0%) of FACS CEA+CT arm subjects diagnosed with recurrence underwent curative-intent treatment compared with a model-predicted

46.3% (95% prediction interval 30.3% to 62.5%). For the minimum arm, however, the observed proportion undergoing curative-intent treatment (8.7%; 95% CI 0.0% to 28.5%) was significantly lower than predicted 29.1% (95% prediction interval 15.1% to 44.4%). For all arms combined, the model's prediction of 34.1% (95% prediction interval 26.5% to 41.3%) matched the observed value exactly (34.1%; 95% CI 25.4% to 43.5%).

Figure 4 shows the observed and predicted five-year overall survival curves for each FACS arm. For each arm, FACS survival curves lay entirely within model prediction intervals for the full five-year period. At five years after detection of recurrence, 27.3% of subjects in the FACS CEA group remained alive compared with a model prediction of 21.8% (95% prediction interval of 9.5% to 35.5%); 26.8% of the CT group remained alive compared with a predicted 21.1% (95% prediction interval 8.8% to 35.3%); 34.5% of the CEA+CT group remained alive compared with a predicted 22.8% (95% prediction interval 10.5% to 35.9%); 17.4% of the minimum surveillance group remained alive compared with a predicted 16.8% (95% prediction interval 5.6% to 30.3%). For all groups combined, 27.0% of subjects were alive at five years after diagnosis of recurrence compared with a prediction of 20.6% (95% prediction interval 14.6% to 26.9%), reflecting slightly pessimistic survival predictions by the model on average.

Figure 5 shows a calibration plot[45, 46]. The 45-degree solid line represents a reference standard of perfect calibration and refinement. The fitted line connecting the data points has y-intercept of 0.08, indicating that predicted proportions alive at each time point were slightly lower than observed on average. A slope of approximately 0.86 reflects the fact that predicted change in the proportion of subjects alive between years 1 and 5 was slightly greater than observed.

DISCUSSION

The long-term objective of this research is to create a model-based tool which leverages the best available clinical evidence to assist colon cancer patients and providers in personalizing post-treatment surveillance based on patient and disease characteristics. The precision survivorship care that such a model can enable has the potential to improve the effectiveness of post-treatment surveillance, and in turn to improve the long-term survival of those treated for colon cancer. Here, we have introduced the CCSuRe model and presented an external validation against independent data from a recent large multi-center randomized trial of surveillance.[2] The model's natural history behavior is based on patient-level data from a multi-center trial involving U.S. colon cancer survivors.[47] This work represents the first independent validation of a colon cancer surveillance model against the results of a modern clinical trial.

The model performed well in external validation (the most robust form of validation[46]) against data obtained from colon cancer patients in the recent FACS trial. Predicted recurrence proportion fell within FACS 95% confidence intervals for the CT, CEA, CT+CEA arms, and for all arms combined, but not for the minimum arm. The model predicted that incidence of recurrence would be similar in each of the follow-up arms since the stage distribution was similar across arms; the mode of surveillance should not in theory affect the probability of recurrence. By contrast the observed number of recurrences differed between arms in the FACS trial, with the fewest being detected in the minimum follow-up arm. This may reflect under-reporting of recurrence in the minimum arm due to less intense follow-up.

The proportions of recurring patients predicted by CCSuRe to undergo curative-intent treatment fell within FACS 95% confidence intervals for the CT, CEA, and CT+CEA arms, and for all arms combined, but again not for the minimum surveillance arm. In the trial, only two patients in the minimum surveillance arm are known to have undergone curative-intent treatment based on available follow-up

data. Accordingly, it is difficult to assess model performance against data from the FACS minimum follow-up arm.

Five-year overall survival for FACS subjects with recurrence in all four arms fell within corresponding model-projected 95% prediction intervals. To assess the calibration and refinement of the model, we created a calibration plot. A slightly positive y-intercept suggests that the model is mildly pessimistic in its survival predictions generally.[46, 48] This may be attributable to improvements in colon cancer treatment and survival in the period between recruitment for the COST Trial (1994 to 1999[37]), whose data was used for model training, and the FACS Trial (2003 to 2009[2]), whose data was used in validation. Overall, the model predicted 5-year survival among subjects with recurrence that was 6.4 percentage points lower than what was reported in the combined FACS trial colon cancer subjects. Over approximately the same timespan (1995-2005) In the U.S., 5-year relative survival for colorectal cancer patients in general improved by 6.5 percentage points.[49] Finally, a calibration plot slope less than 1.0 reflects the fact that predicted change in the proportion of subjects alive between years 1 and 5 was slightly greater than observed in the FACS trial.[46, 50] The slope of 0.86 compares favorably with calibration plot slopes from other external model validations.[45, 50]

We previously developed a simple proof-of-concept model simulating the interaction between recurrence natural history and early detection through surveillance.[31] It was the first published model to account for progression of recurrent disease during diagnostic delay and that considered the full range of possible metastatic sites.[51–54] CCSuRe builds upon our previous model by using individual-level patient data to parameterize functions describing non-constant hazards of state transitions over time. The new model also incorporates covariates accounting for the effect of cancer stage on progression risk and for the effect of age on mortality. CCSuRe employs more realistic algorithms based on empirical data to simulate diagnostic testing following positive surveillance findings—a feature that will enable future cost comparisons between proposed surveillance regimens.

In sensitivity analysis, PET sensitivity was the most influential single parameter on both the likelihood of curative surgery and overall survival. When varied across a 30 percentage point range, the proportion of patients with recurrence who underwent curative treatment in the model varied by 5.0 percentage points, and OS5 varied by 1.5 percentage points. This is not surprising since PET is frequently used in the setting of serially elevated CEA without localized findings as well as to assess resectability when recurrent disease is detected.[20] Model outcomes were also somewhat sensitive to wait times for repeat or follow-up diagnostic tests. This finding suggests that minimizing delays in definitively diagnosing recurrence by minimizing unnecessary wait times for subsequent testing could improve patient outcomes. It also underscores the value that improved testing methodologies leading to fewer equivocal results (e.g. an improved biomarker for CC) could bring.

Limitations

As with many modeling studies, some parameters are derived from disparate sources in the scientific literature. We have examined the implications of mis-specification of such parameters using one-way and multi-way sensitivity analysis. Few inputs changed predicted outcomes to a clinically relevant extent. It should also be noted that morbidity stemming from testing itself (most notably, morbidity stemming from colonoscopy complications) is not included in the model.

The COST trial was the data source used to "train" the disease progression submodel. COST subjects were followed into the early 2000's. Aside from FACS, only one large study which prospectively followed CC survivors undergoing multi-modality post-treatment surveillance has been published since COST.[9] Despite some differences in treatment norms between the COST trial era and today, and despite the differences in settings, it is reassuring from a generalizability perspective that the model projected outcomes which largely approximated what was observed recently among FACS trial colon cancer subjects—albeit with slightly low survival projections across the board. Therapeutic advances

which generally lower recurrence probability after initial treatment should not significantly impact the generalizability of the model for comparing relative benefits of alternative surveillance regimens to patients whose disease recurs. More likely to impact CCSuRe's generalizability would be 1) advances in primary treatment (e.g. surgical technique) which substantially change the anatomic distribution of recurrences, 2) introduction of diagnostic tests with improved performance characteristics, or 3) therapeutic improvements which allow curative treatment of more advanced recurrent disease.

Future Directions

While more intensive surveillance may somewhat improve clinical outcomes, surveillance testing is associated with cost, inconvenience, and potential physical and psychological morbidity. The ability to identify patients who are most likely to benefit from surveillance after surgical cure, and to tailor surveillance to clinical characteristics and preferences of individual patients, promises to increase value from both patient and system perspectives. More extensive and recent longitudinal data describing surveillance testing and outcomes will permit further estimation of existing model parameters for both colon and rectal cancer in the context of more contemporary treatment and diagnostic norms. Further data will also allow estimation of additional risk-modifying covariates. Such covariates could include traditional clinicopathologic risk factors and prognostic or predictive molecular characteristics. The insights gained from this work can inform future risk-stratified surveillance guidelines and clinical decision aids.

Conflict of Interest: The authors declare that they have no conflict of interest.

Dedication

The authors wish to dedicate this manuscript to the memory of Dr. Daniel J. Sargent, whose friendship, collegiality, and contributions to colon cancer research have touched us deeply.

FIGURE LEGENDS

Figure 1 – Sensitivity analysis results – The "tornado diagrams" show the parameters which, when varied across the ranges shown in Table 1, had the greatest impact on the values of a) proportion of recurring patients undergoing curative-intent surgery and b) 5-year overall survival of recurring patients (OS5). PET = positron emission tomography; CEA = carcinoembryonic antigen; CT = computed tomography

Figure 2 - Proportion of patients recurring: model-predicted versus observed in FACS trial. The first of each pair of solid bars represents the mean of 1,000 model runs. Error bars represent 95% prediction intervals for model-predicted outcomes and 95% confidence intervals for observed FACS trial outcomes. CEA = carcinoembryonic antigen; CT = computed tomography; CCSuRe = Colon Cancer Surveillance and Recurrence Model; FACS = Follow-up After Colorectal Surgery Trial

Figure 3 – Proportion of recurring patients who were treated with curative intent: model-predicted versus observed in FACS trial. The first of each pair of solid bars represents the mean of 1,000 model runs. Error bars represent 95% prediction intervals for model-predicted outcomes and 95% confidence intervals for observed FACS trial outcomes. CEA = carcinoembryonic antigen; CT = computed tomography; CCSuRe = Colon Cancer Surveillance and Recurrence Model; FACS = Follow-up After Colorectal Surgery Trial

Figure 4 (a-d) – Overall survival following recurrence diagnosis: model-predicted versus observed in FACS trial. a) CEA arm, b) CT arm, c) CEA+CT arm, d) minimum arm – The solid line in each graph represents the mean of 1,000 model runs. The shaded regions represent 95% predictions intervals. CCSuRe = Colon Cancer Surveillance and Recurrence Model; FACS = Follow-up After Colorectal Surgery Trial

Figure 5 – Calibration plot comparing overall survival of FACS trial recurrers to mean model-predicted overall survival at years one through five – The y-intercept and slope of the line fitted to these points were used to assess model calibration (i.e. correspondence between the overall levels of predicted and observed values) and refinement (i.e. degree to which model estimates span a range similar to that

spanned by actual observations) , respectively.[45, 46] The 45-degree solid line represents a reference standard of perfect calibration and refinement.

TABLES

		Parameter estimate	Range used in sensitivity analysis	Source
Surveillance and diagnostic test characteristics				
	CEA sensitivity	0.64	0.49 to 0.79	[55]
	CEA specificity	0.90	0.75 to 1.00	[55]
	CT sensitivity – chest	0.94	0.79 to 1.00	[56]
	CT specificity – chest	0.96	0.81 to 1.00	[56]
	Ct sensitivity – liver	0.83	0.68 to 0.98	[57]
	CT specificity – liver	0.59	0.44 to 0.74	[57]
	CT sensitivity – abdomen	0.73	0.58 to 0.88	[58] [56] [59]
	CT specificity - abdomen	0.98	0.83 to 1.00	[56]
	Colonoscopy sensitivity	0.95	0.78 to 1.00	[60] [61]
	Colonoscopy specificity	1.00	0.85 to 1.00	[60] [61]
	Liver MRI sensitivity	0.86	0.71 to 1.00	[57]
	Liver MRI specificity	0.87	0.72 to 1.00	[57]
	PET sensitivity	0.87	0.72 to 1.00	[62]
	PET specificity	0.96	0.81 to 1.00	[62]
Probability of equivocal test results given that recurrence is present				
	Probability of equivocal CEA	0.55	0.40 to 0.70	[a]
	Probability of equivocal CT	0.136	0.00 to 0.286	[a]
Intervals for repeat or follow-up diagnostic testing				
	For repeat CEA after equivocal result	6 weeks	3 to 12 weeks	[a,b]
	For repeat CT after equivocal CT	4 weeks	2 to 8 weeks	[a,b]
	For follow-up CT scan after positive CEA	2 weeks	1 to 4 weeks	[a,b]
	For follow-up liver MRI after positive CT	1 week	0 to 2 weeks	[a,b]
	For follow-up PET scan	1 week	0 to 2 weeks	*, †

[a] Internal dataset available upon request

[b] Expert opinion

Table 1 – Utilization submodel parameters – Specific testing algorithms are shown in Appendix 3. CEA = carcinoembryonic antigen; CT = computed tomography; MRI = magnetic resonance imaging; PET = positron emission tomography

		CEA Surveillance (n=217)	CT Surveillance (n=202)	CEA + CT Surveillance (n=208)	Minimum Surveillance (n=214)
Subject Characteristics (N=841)					
	Mean age at randomization	69.4	69.0	70.6	69.3
	Dukes Stage of primary colon cancer (A/B/C)	14.3% / 55.8% / 30.0%	19.8% / 47.5% / 32.7%	15.4% / 56.3% / 28.4%	17.8% / 50.0% / 32.2%
Surveillance Regimens					
	CEA frequency	Every 3 months for 2 years, then every 6 months until year 5	None	Every 3 months for 2 years, then every 6 months until year 5	None
	CT frequency	At 12-18 months	Every 6 months for 2 years, then annually until year 5	Every 6 months for 2 years, then annually until year 5	At 12-18 months
	Colonoscopy frequency	None	At 2 years	At 2 years	None

Table 2 – Sample characteristics and surveillance regimens from FACS trial colon cancer patients. CEA = carcinoembryonic antigen; CT = computed tomography

APPENDIX 1 – Detailed description of disease progression submodel

1. Multi-state Markov models

We use a multi-state non-homogenous Markov model to model underlying disease progression; this type of model has been used previously to model the progression of cancer and other chronic illnesses for which transitions between disease states cannot be observed directly.[32–36] Let $X(t)$ be a continuous time Markov process which can take on n progressively severe states, denoted by$\{1,2, \dots n\}$, with n being an absorbing state. For any two states i and j such that$i \neq j$, the transition intensity $\lambda_{ij}(t)$ is the rate at which transitions occur from state i to state j at time t, that is,

$$\lambda_{ij}(t) = \lim_{h \to 0} \frac{P(X(t+h) = j \,|X(t) = i)}{h}.$$

The function $\lambda_{ij}(t)$ can also be thought of as the hazard of transitioning from state i to state j at time t. Note further that if $i = j$, then

$$\lambda_{ii}(t) = -\sum_{\substack{j=1 \\ i \neq j}}^{n} \lambda_{ij}(t).$$

Although frequently it is assumed that the hazard rates $\lambda_{ij}(t)$ are constant with respect to time, in this model, we make use of a non-homogenous Markov process, assuming that some of the transition intensities vary with time. Furthermore, we define the transition probabilities for this process by

$$p_{ij}(s,t) = P\,(X(t) = j \mid X(s) = i),$$

for any two states *i* and *j* and times 0 < *s* < *t*. Given the transition intensities for a process, the corresponding transition probabilities are found by solving the Kolmogorov forward equations.[63]

2. Application to colon cancer and selection of hazard functions

For our application, we assume that $X(t)$ represents an individual's current disease state at time $t \geq 0$ after initial surgery. Here $X(t)$ can take on 4 possible disease states, listed below:

(1) No known recurrence
(2) Detectable and resectable recurrence
(3) Detectable recurrence that is not resectable
(4) Death

Figure A1 shows a diagram of these states, along with all of the possible transitions in our model. We assume each of the transition intensities $\lambda_{ij}(t)$ in our model takes on a parametric form, which we describe in detail below.

Consider first the transitions from state 1 to state 2 and from state 1 to state 3. Note that both of these transitions represent the event of a recurrence; however, a transition from 1 to 2 represents a recurrence that could potentially be resectable if it was detected early enough, while a transition from 1 to 3 is a recurrence that is never resectable, regardless of when it is detected. This determination of whether a recurrence could ever be resectable is made based on the site(s) of the recurrence as described in the main text. We assume that the transition intensities $\lambda_{12}(t)$ and $\lambda_{13}(t)$ are equal.

We know from empirical hazard plots for time to recurrence detection from the Clinical Outcomes of Surgical therapy (COST) Trial[37] that the hazard of recurrence is non-constant over time; it increases initially in the time after surgery, reaches a maximum, and then declines. Perez-Ocon et al[64] suggest using a piecewise Weibull hazard to model a similar pattern in the timing of breast cancer recurrence. Note that the Weibull hazard function has the form

$$\lambda(t) = \frac{\gamma}{\alpha}\left(\frac{t}{\alpha}\right)^{\gamma-1}$$

for fixed shape and scale parameters γ and α. Notice that for $\gamma < 1$, the hazard is monotonically decreasing, while for $\gamma > 1$, it is monotonically increasing. (For $\gamma = 1$, it is constant).

In addition to being non-constant with respect to time, the hazard of recurrence also differs based on the stage of a patient's primary cancer; thus we include covariates for stage in $\lambda_{12}(t)$ using the proportional hazards assumption.

To write an expression for $\lambda_{12}(t)$ using this approach, we first define the function $\delta_{12}(t)$,

$$\delta_{12}(t) = \begin{cases} \frac{\gamma_{12}^{(1)}}{\alpha_{12}^{(1)}}\left(\frac{t}{\alpha_{12}^{(1)}}\right)^{\gamma_{12}^{(1)}-1}, & 0 \le t \le T_{break} \\ \frac{\gamma_{12}^{(2)}}{\alpha_{12}^{(2)}}\left(\frac{t - T_{break}}{\alpha_{12}^{(2)}}\right)^{\gamma_{12}^{(2)}-1}, & t > T_{break} \end{cases}$$

for a fixed value of T_{break} that we determine empirically from the data. Note that T_{break} is the time after initial surgery when the hazard of recurrence is at its maximum. Then the expression for $\lambda_{12}(t)$ is given by

$$\lambda_{12}(t) = \exp(\beta_{12}^{s2} I_{s2} + \beta_{12}^{s3} I_{s3})\delta_{12}(t)$$

where I_{s2} is an indicator variable equal to 1 for stage 2 patients and 0 otherwise; similarly, I_{s3} is equal to 1 only for stage 3 patients.

Next, we consider the transition from state 2 to state 3. The later a patient recurs after their initial surgery, the longer the time they will spend in state 2.[38][39][40][41] Thus, we model this transition with a declining hazard; as before, we choose a Weibull hazard, so that

$$\lambda_{23}(t) = \frac{\gamma_{23}}{\alpha_{23}}\left(\frac{t}{\alpha_{23}}\right)^{\gamma_{23}-1},$$

where t is the time since initial surgery.

Consider next the transitions from states 1 and 2 to state 4. These transitions both represent death from other causes; thus we assume that $\lambda_{14}(t) = \lambda_{24}(t)$. Moreover, we use a constant hazard with a covariate for age to model these transitions,

$$\lambda_{14}(t) = \exp(\beta_{14}(age))\,\alpha_{14},$$

where α_{14} is a constant rate. Lastly, we consider the transition from state 3 to state 4. A patient can make this transition by dying from other causes or by dying from cancer. Thus, we assume that $\lambda_{34}(t)$ has the form,

$$\lambda_{34}(t) = \alpha_{34} + \exp(\beta_{14}(age))\,\alpha_{14},$$

where α_{34} is the (assumed) constant hazard of death from cancer. We then must estimate the set of natural history parameters $\theta = \left\{\alpha_{12}^{(1)},\ \gamma_{12}^{(1)},\ \alpha_{12}^{(2)},\ \gamma_{12}^{(2)},\ \beta_{12}^{s2},\ \beta_{12}^{s3},\ \alpha_{23},\ \gamma_{23},\ \alpha_{14},\ \beta_{14},\ \alpha_{34}\right\}$ from the COST Trial data.

Note that in Castelli et al[65], a multi-state semi-Markov model was used to model the natural history of colorectal cancer recurrence; this model also employed Weibull hazards for the transitions between states. However, this model utilized a single state for recurrence and did not differentiate between resectable and unresectable recurrences. Furthermore, this model assumed a single Weibull hazard for the transition between the no recurrence state and recurrence state, instead of the piecewise hazard that we use here.

3. Maximum likelihood estimation and misclassification

Now suppose that we have a set of *N* observations from the process; the *i*th observation consists of a set of m_i observation times $t_k, 1 \leq k \leq m_i$, and corresponding observed states $o_k,\ 1 \leq k \leq m_i$. Note that we will differentiate here between the observed state of the process o_k at time t_k and the actual state, which we will denote by x_k. We make this distinction because in our application, the true disease state can only be observed indirectly through the results of a diagnostic test. Thus, we only have information about the observed disease state at a given time, which may not correspond to the true disease state at that time, and we must account for this potential error in our parameter estimation.

The problem of parameter estimation in the presence of potential misclassification has been addressed before (see for instance [33, 34, 66]). Briefly, let $O(t)$ denote the observed disease state at time t; recall that we denote the true disease state at time t by $X(t)$. The value of $O(t)$ is related to the value of $X(t)$ via a series of misclassification probabilities, defined as $P(O(t) = i\ |X(t) = j)$, for any two states i and j.

The likelihood function which accounts for misclassification is constructed in the following manner[33]: assuming that the observation at t_1 is made without error, the contribution to the likelihood for one set of observations is given by

$$\ell_i(\theta) = \sum_{x_2, x_3 \ldots x_{mi}} \prod_{k=2}^{m_i} P(O(t_k) = o_k | X(t_k) = x_k)\, p_{x_{k-1} x_k}(t_{k-1}, t_k),$$

where the summation is taken over all possible values of the true underlying states $x_2, x_3, \ldots x_{m_i}$. The misclassification probabilities $P(O(t_k) = o_k \,|X(t_k) = x_k)$ in our application are given by the sensitivities and specificities of the surveillance tests used to make the observations. The total likelihood for N observations is then given by

$$L(\theta) = \prod_{i=1}^{N} \ell_i(\theta).$$

We then estimate the set of the natural history parameters θ that maximizes $\log L(\theta)$ using a quasi-Newton algorithm.[67]

A more detailed description of maximum likelihood estimation of model parameters is provided in Appendix 2.

4. Symptom parameter estimation

In addition to the disease progression parameters, we also require an estimate of the rate at which patients who recur develop symptoms. We assume that patients become eligible to develop symptoms when they make the transition from state 1 to state 2 (or from state 1 to state 3, for those who do not pass through state 2). The development of symptoms does not differ based on whether a patient is in state 2 or state 3. We assume a constant hazard for the development of symptoms, that is,

$$\lambda_{sym}(t) = \alpha_{sym}.$$

Our method for estimating this parameter requires an estimate of the disease progression parameters θ; thus, we estimate the rate of symptom development separately after we have computed the maximum likelihood estimates of the other natural history parameters.

Because the COST trial data did not include descriptions of symptom status at presentation, we used a separate dataset (available upon request) based on 62 curatively resected colon cancer patients at our own institution to develop the hazard rate estimate. This internal dataset contained details on all relevant healthcare encounters following initial CC treatment and symptom status at each.

Suppose that for each individual in a set of N patients with recurrences, we have the time when they were last definitively known to be without recurrence (which we denote by $t_{lastnorm}$) as well as their time of detection of recurrence (denoted t_{det}); in addition, it is known whether or not the patient was symptomatic at the time of detection. For patients who were symptomatic at time of detection, we also know the time symptoms began (denoted t_{sym}; note that $t_{sym} \leq t_{det}$).

Denote by $p_{1(nosym)}(s,t)$ the probability that an individual who recurred between times s and t did not develop symptoms. Note that $p_{1(nosym)}(s,t)$ is a function of α_{sym} as well as the disease progression parameters. Similarly, define $p_{(nosym)(sym)}(s,t)$ to be the probability that a patient with recurrence develops symptoms between times s and t. This probability only depends on the value α_{sym}. We compute the maximum likelihood estimate of α_{sym}. For an individual who recurred but had not developed symptoms by their time of detection, the contribution to the likelihood is given by

$$\ell_i(\alpha_{sym}) = p_{1(nosym)}(t_{lastnorm}, t_{det}).$$

The contribution to the likelihood by an individual who recurred and had developed symptoms before they were detected is given by

$$\ell_i(\alpha_{sym}) = p_{1(nosym)}(t_{lastnorm}, t_{sym} - d)p_{(nosym)(sym)}(t_{sym} - d, t_{sym}),$$

where d represents one day. The total likelihood is then given by

$$L(\alpha_{sym}) = \prod_{i=1}^{N} \ell_i(\alpha_{sym}).$$

We find the value of α_{sym} that maximizes $L(\alpha_{sym})$ using a quasi-Newton algorithm.

5. Survival for recurrers with curative surgery

Recall that when we computed the natural history parameters, we computed an estimate for the rate of transition from state 3 to state 4. Thus, we already have an estimate for the rate of survival for recurrers who were detected after their recurrence had become unresectable. However, we have not considered survival for recurrers who undergo curative-intent surgery (i.e. who are detected in state 2).

For patients who underwent curative surgery for recurrence, we have time of detection as well as time of death or the time they were last known to be alive. We fit an exponential survival model to this data, giving us the rate of survival post-recurrence diagnosis for patients who were detected while their recurrence was still potentially resectable.

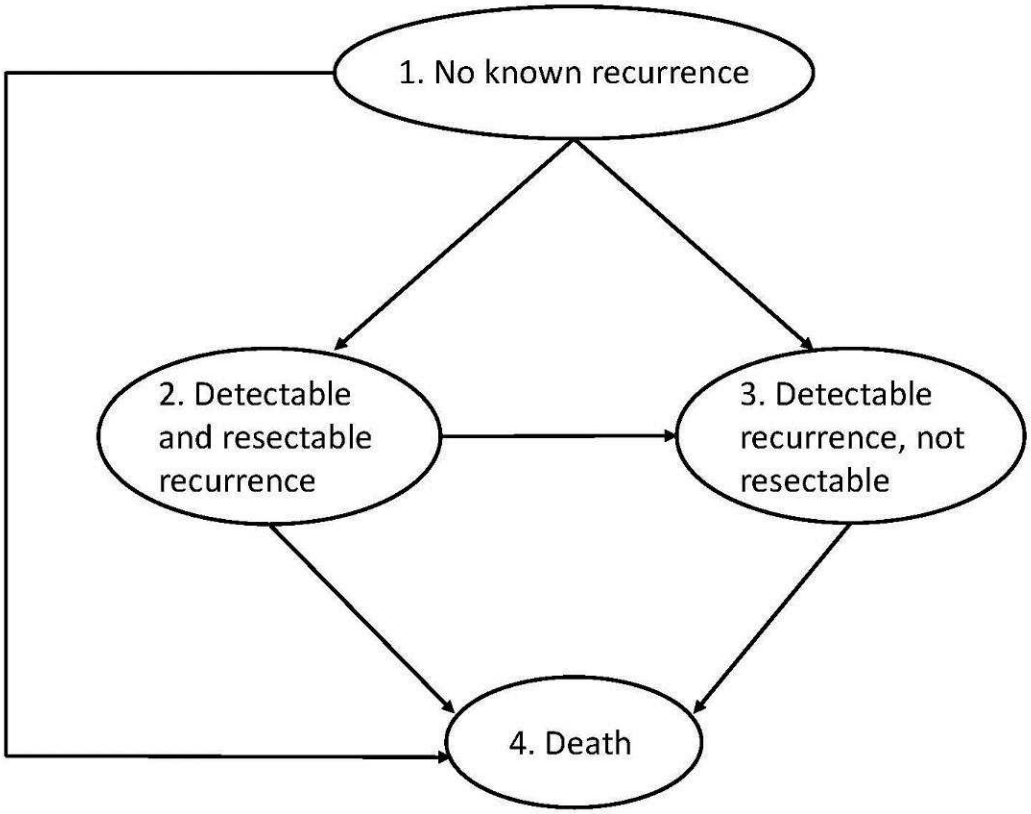


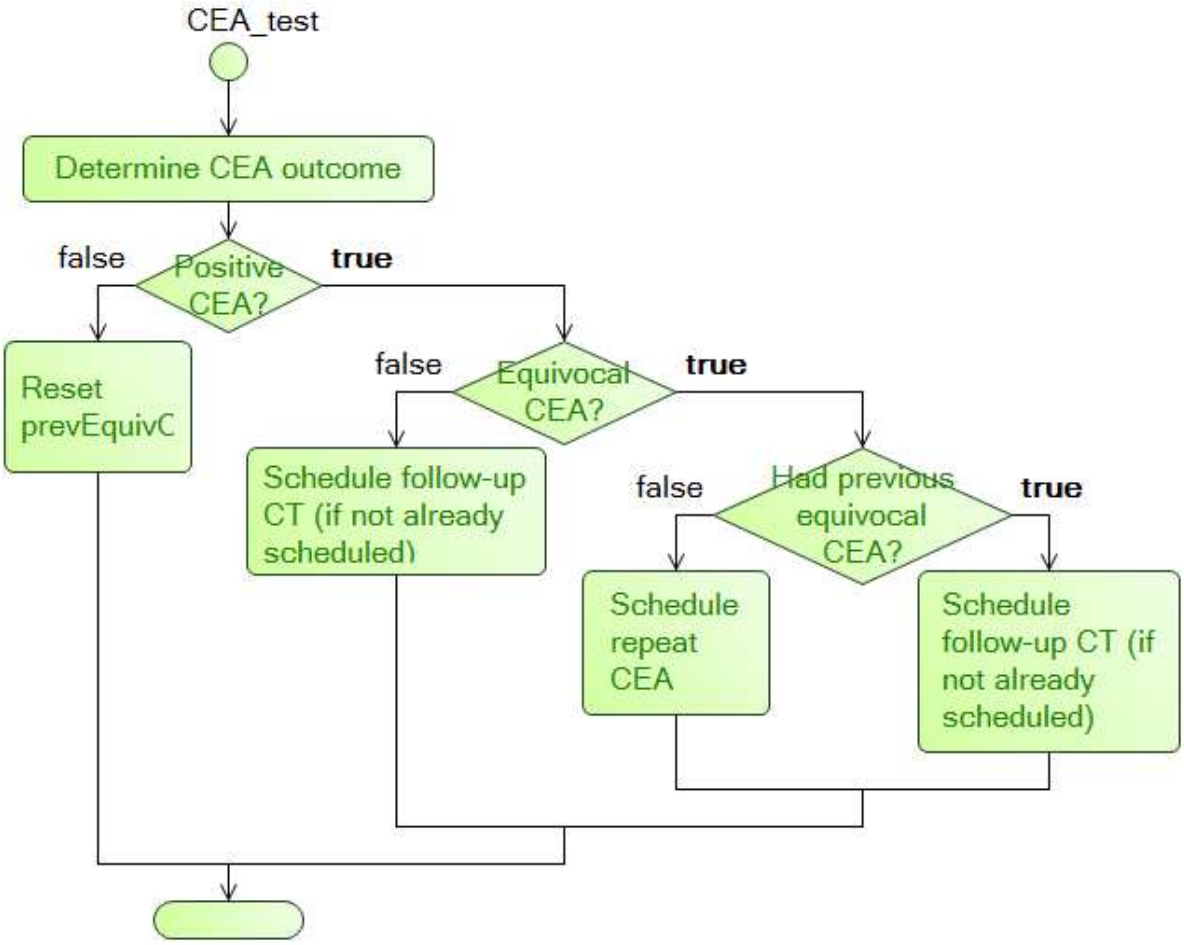
Figure A1. State transition diagram for disease states of the model; the arrows indicate possible transitions.

INSERT APPENDIX 2

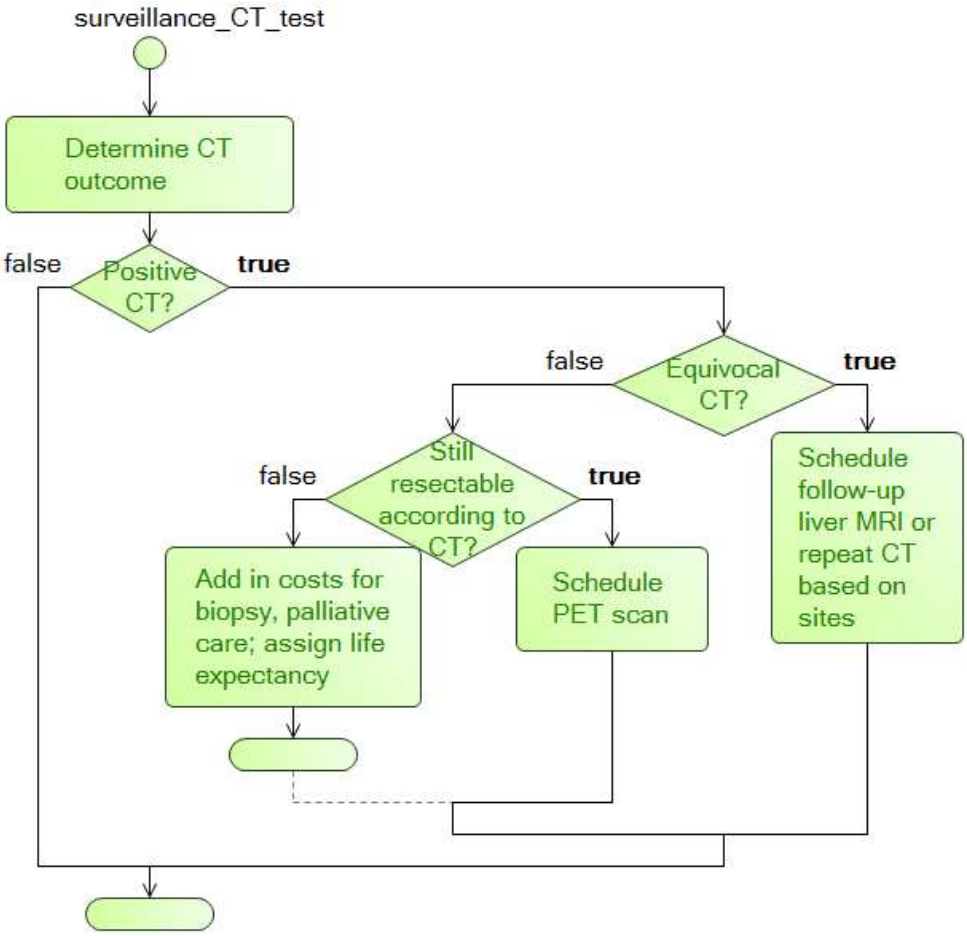
APPENDIX 3 – Utilization submodel algorithms

Note that empty ovals represent exit points for an algorithm whereby an individual returns to "waiting" for the next event. CEA = carcinoembryonic antigen; CT = computed tomography; PET = positron emission tomography; MRI = magnetic resonance imaging

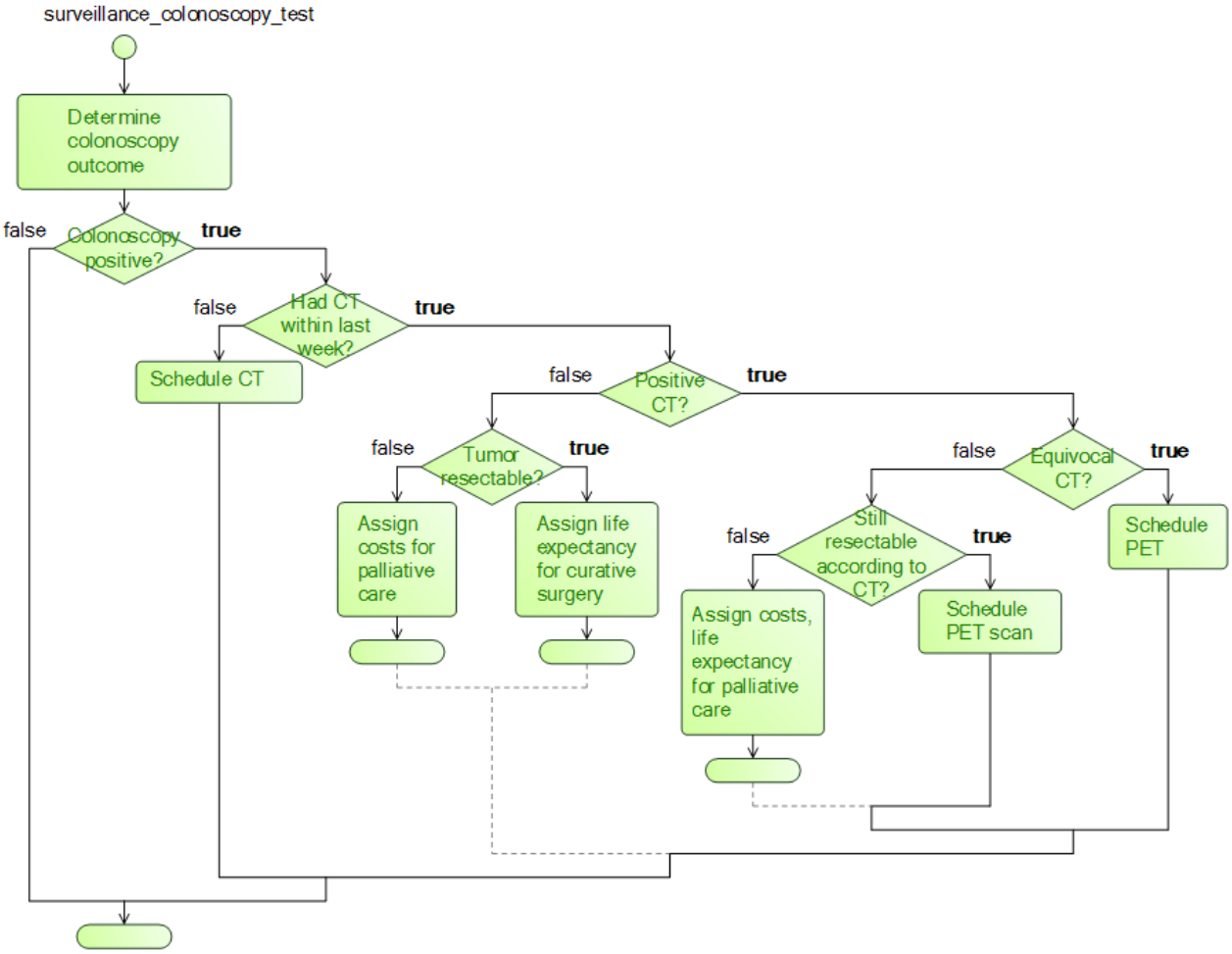
A) Surveillance CEA Testing



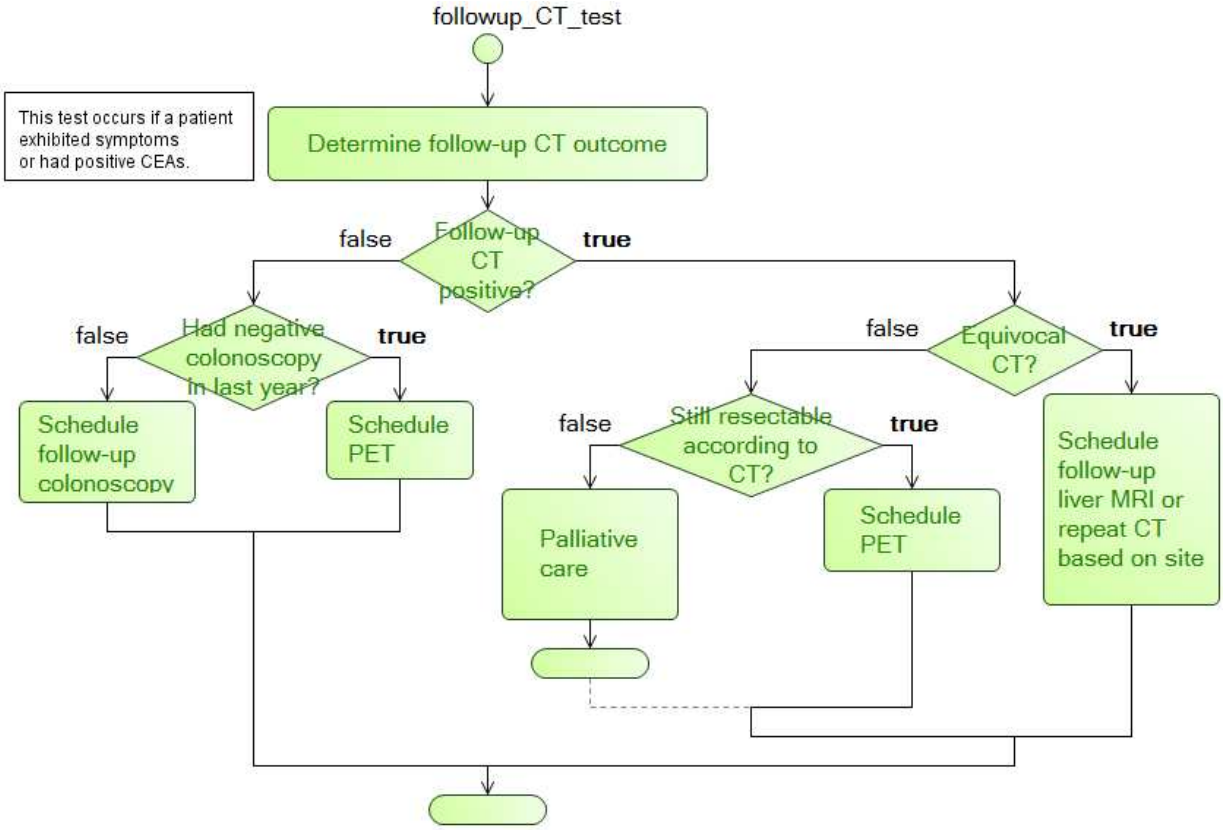
B) Surveillance CT



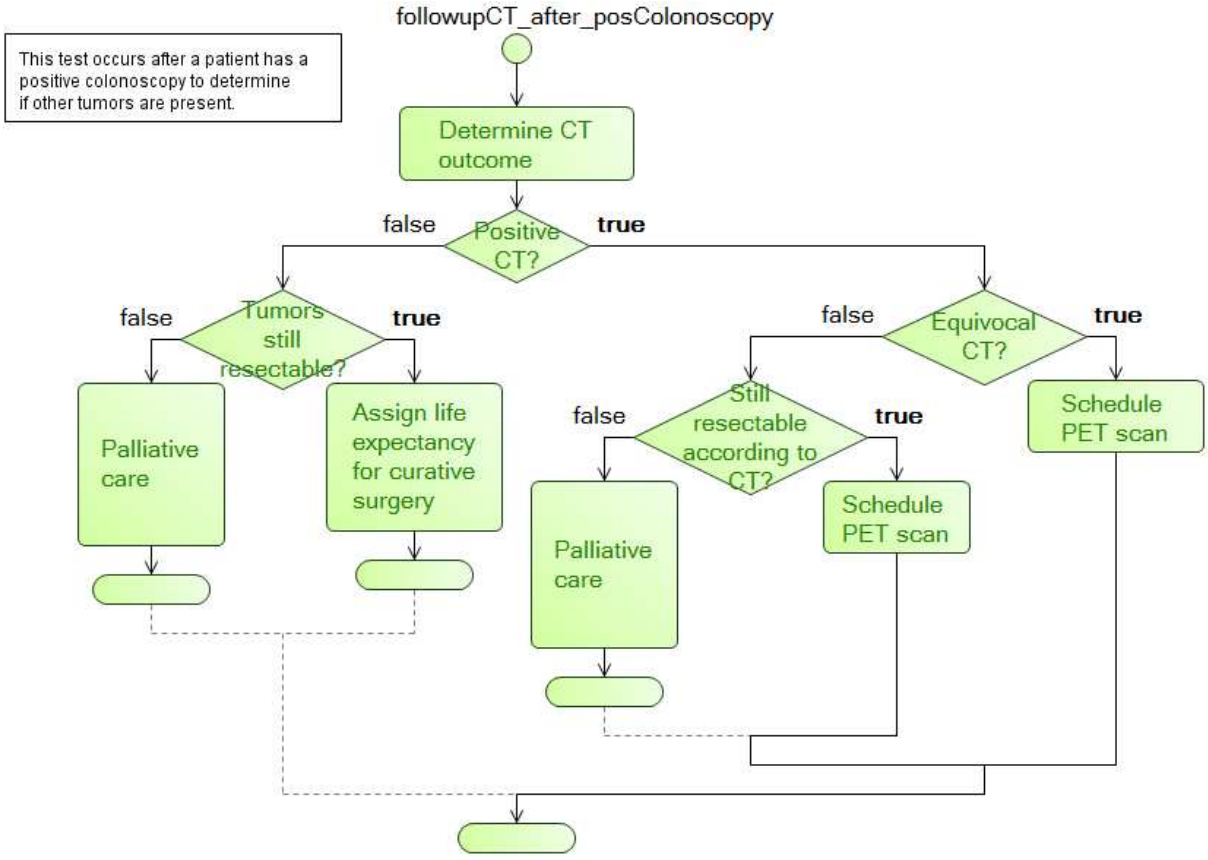
C) Surveillance Colonoscopy



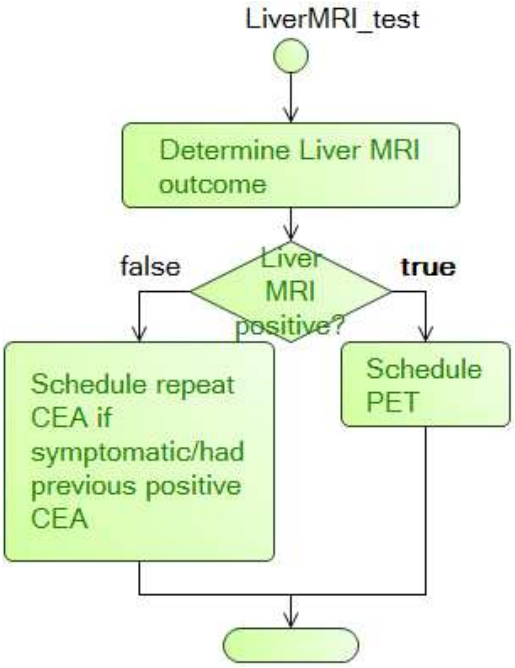
D) Follow-up CT



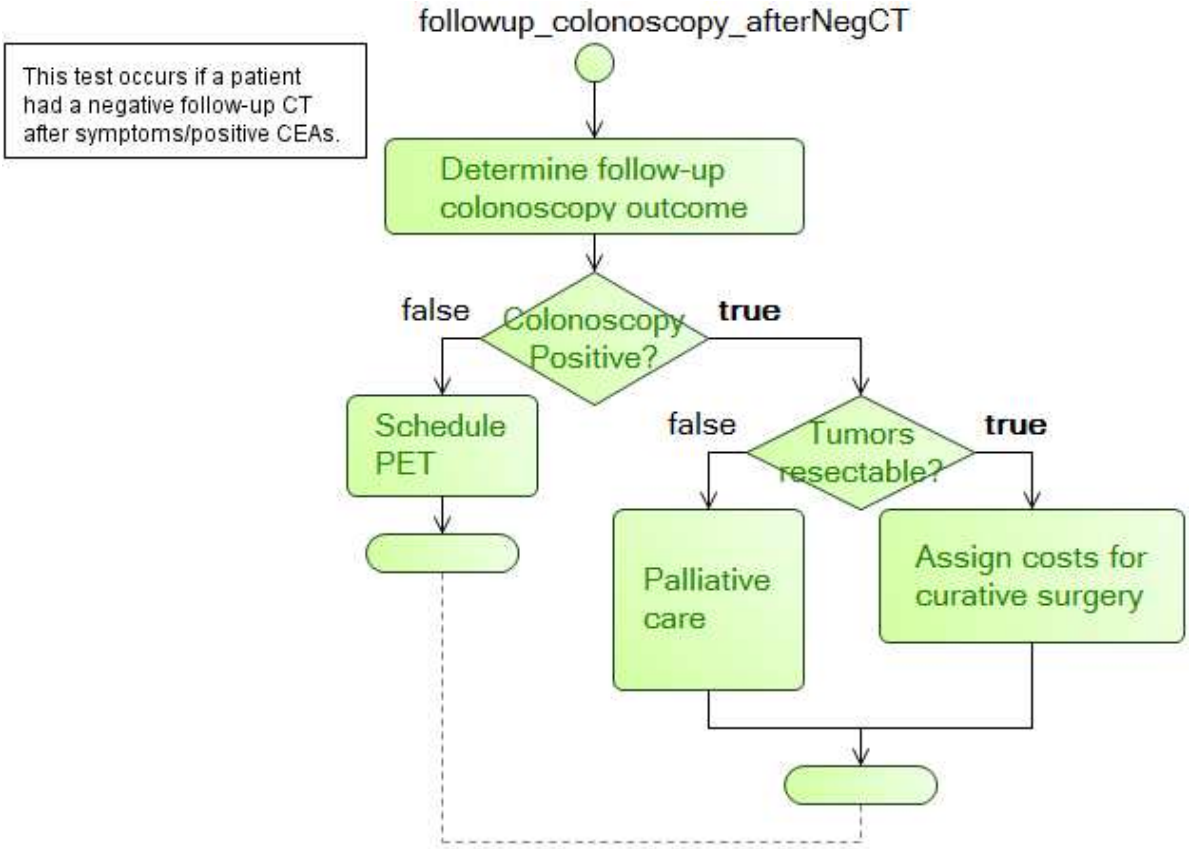
E) CT after Positive Colonoscopy



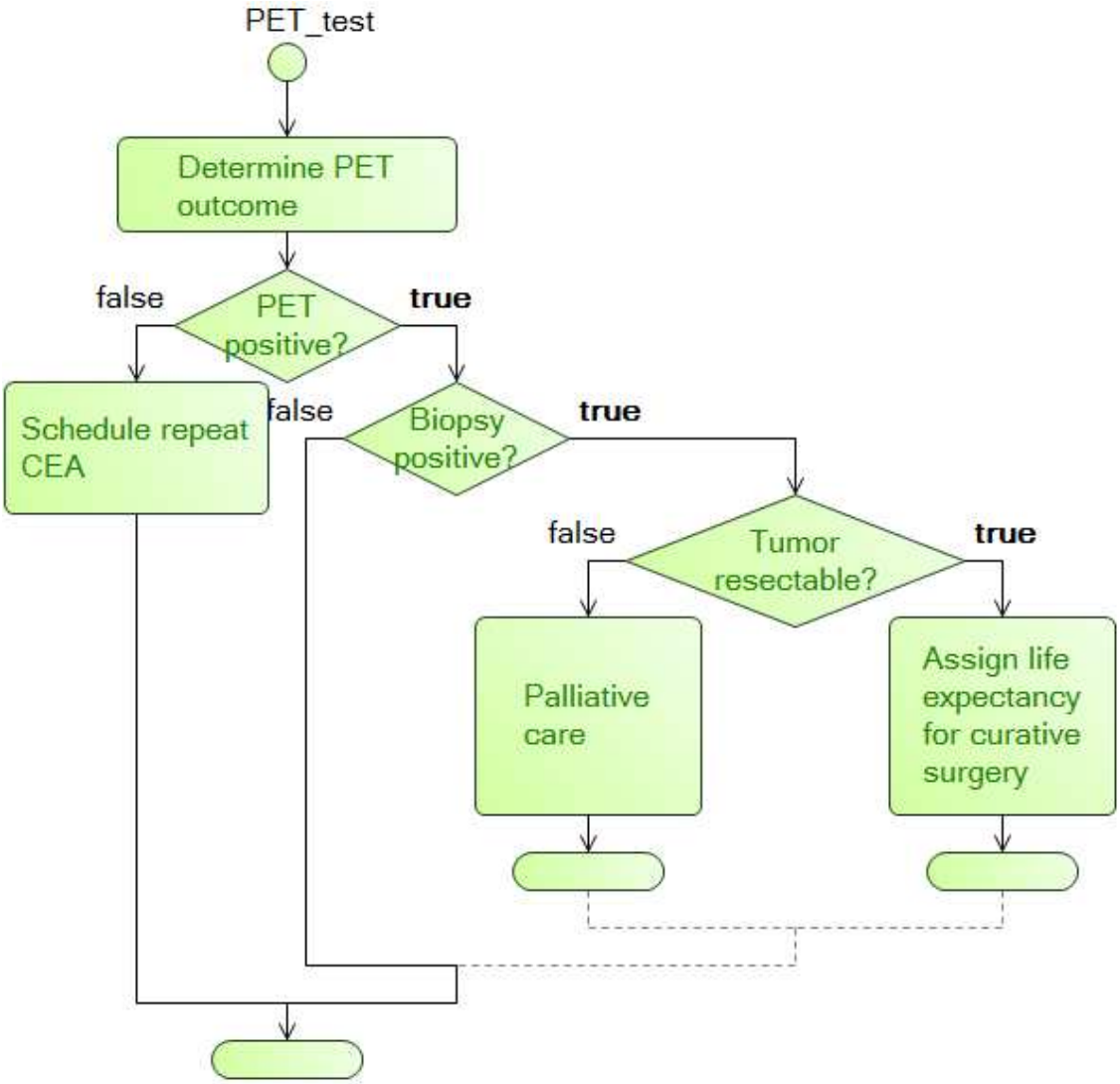
F) Follow-up Liver MRI after CT showing possible liver metastases



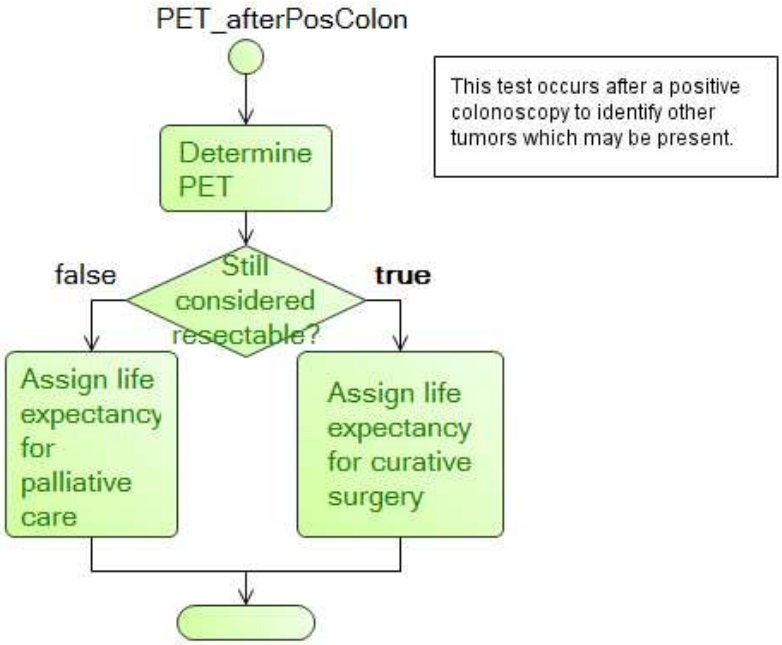
G) Colonoscopy after Negative CT



H) Follow-up PET



I) PET after Positive Colonoscopy



REFERENCES

1. Siegel RL, Miller KD, Jemal A (2016) Cancer statistics, 2016. CA Cancer J Clin 66:7–30 . doi: 10.3322/caac.21332

2. Primrose JN, Perera R, Gray A, et al (2014) Effect of 3 to 5 years of scheduled CEA and CT follow-up to detect recurrence of colorectal cancer: the FACS randomized clinical trial. JAMA 311:263–70 . doi: 10.1001/jama.2013.285718

3. Cunningham D, Atkin W, Lenz HJ, et al (2010) Colorectal cancer. Lancet 375:1030–1047 . doi: 10.1016/S0140-6736(10)60353-4

4. Formica V, Roselli M (2015) Targeted therapy in first line treatment of RAS wild type colorectal cancer. World J Gastroenterol 21:2871–2874 . doi: 10.3748/wjg.v21.i10.2871

5. Renehan AG, Egger M, Saunders MP, O'Dwyer ST (2002) Impact on survival of intensive follow up after curative resection for colorectal cancer: systematic review and meta-analysis of randomised trials. BMJ 324:813 . doi: 10.1136/bmj.324.7341.813

6. Jeffery M, Hickey BE, Hider PN (2007) Follow-up strategies for patients treated for non-metastatic colorectal cancer. Cochrane Database Syst Rev CD002200 . doi: 10.1002/14651858.CD002200.pub2

7. Mokhles S, Macbeth F, Farewell V, et al (2016) Meta-analysis of colorectal cancer follow-up after potentially curative resection. Br J Surg 103:1259–1268 . doi: 10.1002/bjs.10233

8. Rodriguez-Moranta F, Salo J, Arcusa A, et al (2006) Postoperative surveillance in patients with colorectal cancer who have undergone curative resection: a prospective, multicenter, randomized, controlled trial. J Clin Oncol 24:386–393 . doi: 10.1200/JCO.2005.02.0826

9. Rosati G, Ambrosini G, Barni S, et al (2016) A randomized trial of intensive versus minimal surveillance of patients with resected Dukes B2-C colorectal carcinoma. Ann Oncol 27:274–280 . doi: 10.1093/annonc/mdv541

10. Secco GB, Fardelli R, Rovida S, et al (2000) Is intensive follow-up really able to improve prognosis of patients with local recurrence after curative surgery for rectal cancer? Ann Surg Oncol 7:32–37 . doi: 10.1007/s10434-000-0032-y

11. Pietra N, Sarli L, Costi R, et al (1998) Role of follow-up in management of local recurrences of colorectal cancer: a prospective, randomized study. Dis Colon Rectum 41:1127–1133 . doi: 10.1007/BF02239434

12. Schoemaker D, Black R, Giles L, Toouli J (1998) Yearly colonoscopy, liver CT, and chest radiography do not influence 5-year survival of colorectal cancer patients. Gastroenterology 114:7–14 . doi: 10.1016/S0016-5085(98)70626-2

13. Makela JT, Laitinen SO, Kairaluoma MI (1995) Five-year follow-up after radical surgery for colorectal cancer. Results of a prospective randomized trial. Arch Surg 130:1062–1067 . doi: 10.1001/archsurg.1995.01430100040009

14. Verberne CJ, Zhan Z, Van Den Heuvel E, et al (2015) Intensified follow-up in colorectal cancer patients using frequent Carcino-Embryonic Antigen (CEA) measurements and CEA-triggered imaging: Results of the randomized " CEAwatch " trial. doi:

10.1016/j.ejso.2015.06.008

15. Ohlsson B, Breland U, Ekberg H, et al (1995) Follow-up after curative surgery for colorectal carcinoma. Randomized comparison with no follow-up. Dis Colon Rectum 38:619–626 . doi: 10.1007/BF02054122

16. Labianca R, Nordlinger B, Beretta GD, et al (2013) Early colon cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 24: . doi: 10.1093/annonc/mdt354

17. Glynne-Jones R, Wyrwicz L, Tiret E, et al (2017) Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 28:iv22-iv40 . doi: 10.1093/annonc/mdx224

18. NICE. National Institute for Health and Care Excellence (2011) Colorectal cancer: diagnosis and management Clinical guideline. Natl Inst Heal Care Excell 1–48

19. Rex DK, Kahi CJ, Levin B, et al (2006) Guidelines for colonoscopy surveillance after cancer resection: a consensus update by the American Cancer Society and the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 130:1865–1871 . doi: 10.1053/j.gastro.2006.03.013

20. National Comprehensive Cancer Network (2016) NCCN Clinical Practice Guidelines in Oncology: Colon Cancer. http://www.nccn.org/professionals/physician_gls/pdf/colon.pdf

21. National Comprehensive Cancer Network (2014) NCCN Clinical Practice Guidelines in Oncology: Rectal Cancer. http://www.nccn.org/professionals/physician_gls/pdf/rectal.pdf

22. Meyerhardt JA, Mangu PB, Flynn PJ, et al (2013) Follow-up care, surveillance protocol, and secondary prevention measures for survivors of colorectal cancer: American society of clinical oncology clinical practice guideline endorsement. J Clin Oncol 31:4465–4470 . doi: 10.1200/JCO.2013.50.7442

23. Weiser MR, Landmann RG, Kattan MW, et al (2008) Individualized prediction of colon cancer recurrence using a nomogram. J Clin Oncol 26:380–385 . doi: 10.1200/JCO.2007.14.1291

24. Huh JW, Lee JH, Kim HR, Kim YJ (2013) Prognostic significance of lymphovascular or perineural invasion in patients with locally advanced colorectal cancer. Am J Surg 206:758–763 . doi: 10.1016/j.amjsurg.2013.02.010

25. Lanza G, Gafà R, Maestri I, et al (2002) Immunohistochemical pattern of MLH1/MSH2 expression is related to clinical and pathological features in colorectal adenocarcinomas with microsatellite instability. Mod Pathol 15:741–9 . doi: 10.1097/01.MP.0000018979.68686.B2

26. Belt EJT, Stockmann HBAC, Abis GSA, et al (2012) Peri-Operative Bowel Perforation in Early Stage Colon Cancer is Associated with an Adverse Oncological Outcome. J Gastrointest Surg 16:2260–2266 . doi: 10.1007/s11605-012-2053-9

27. Krarup P-M, Nordholm-Carstensen A, Jorgensen LN, Harling H (2014) Anastomotic Leak Increases Distant Recurrence and Long-Term Mortality After Curative Resection for Colonic Cancer. Ann Surg 259:930–938 . doi: 10.1097/SLA.0b013e3182a6f2fc

28. Vargas GM, Sheffield KM, Parmar AD, et al (2013) Physician follow-up and observation of guidelines in the post treatment surveillance of colorectal cancer. Surg (United States) 154:244–255 . doi: 10.1016/j.surg.2013.04.013

29. van der Stok EP, Spaander MCW, Grunhagen DJ, et al (2016) Surveillance after curative treatment

for colorectal cancer. Nat Rev Clin Oncol. doi: 10.1038/nrclinonc.2016.199

30. Gilardoni E, Bernasconi DP, Poli S, et al (2015) Surveillance for early stages of colon cancer: potentials for optimizing follow-up protocols. World J Surg Oncol 13:260 . doi: 10.1186/s12957-015-0674-7

31. Rose J, Augestad KM, Kong CY, et al (2014) A simulation model of colorectal cancer surveillance and recurrence. BMC Med Inform Decis Mak 14:29 . doi: 10.1186/1472-6947-14-29

32. Uhry Z, Hedelin G, Colonna M, et al (2010) Multi-state Markov models in cancer screening evaluation: a brief review and case study. Stat Methods Med Res 19:463–486

33. Jackson CH, Sharples LD, Thompson SG, Duffy SW (2003) Multistate Markov models for disease progression with classification error. Stat 52:193–209

34. Chen H, Duffy S, Tabar L (1996) A Markov chain method to estimate the tumor progression rate from preclinical to clinical phase, sensitivity and postiive predictive value for mammography in breast cancer screening. Stat 45:307–317

35. Putter H, van der Hage J, de Bock GH, et al (2006) Estimation and Prediction in a Multi-State Model for Breast Cancer. Biometrical J 48:366–380

36. Dancourt V, Quantin C, Abrahamowicz M, et al (2004) Modeling recurrence in colorectal cancer. J Clin Epidemiol 57:243–251

37. Clinical T, Therapy S, Group S (2004) A Comparison of Laparoscopically Assisted and Open Colectomy for Colon Cancer. N Engl J Med 350:2050–2059 . doi: 10.1056/NEJMoa032651

38. O'Connell MJ, Campbell ME, Goldberg RM, et al (2008) Survival following recurrence in stage II and III colon cancer: findings from the ACCENT data set. J Clin Oncol 26:2336–2341 . doi: 10.1200/JCO.2007.15.8261

39. Hughes KS, Rosenstein RB, Songhorabodi S, et al (1988) Resection of the liver for colorectal carcinoma metastases. A multi-institutional study of long-term survivors. Dis Colon Rectum 31:1–4 . doi: 10.1007/BF02552560

40. Olak J, Wexler MJ, Rodriguez J, McLean AP (1986) Hepatic resection for metastatic disease. Can J Surg 29:435–439

41. Petrelli NJ, Nambisan RN, Herrera L, Mittelman A (1985) Hepatic resection for isolated metastasis from colorectal carcinoma. Am J Surg 149:205–209 . doi: 10.1016/S0002-9610(85)80065-9

42. Surveillance, Epidemiology and ER (SEER) P (2017) SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2016 Sub (1973-2014 varying) - Linked To County Attributes - Total U.S., 1969-2015 Counties, National Cancer Institute, DCCPS

43. American Joint Committee on Cancer (2010) AJCC Cancer Staging Handbook

44. Lawless JF, Fredette M (2005) Frequentist prediction intervals and predictive distributions. Biometrika 92:529–542 . doi: 10.1093/biomet/92.3.529

45. Steyerberg EW, Vickers AJ, Cook NR, et al (2010) Assessing the performance of prediction models : A framework for some traditional and novel measures. Epidemiology 21:128–138 . doi: 10.1097/EDE.0b013e3181c30fb2.Assessing

46. Miller ME, Langefeld CD, Tierney WM, et al (1993) Validation of probabilistic predictions. Med Decis Making 13:49–58 . doi: 10.1177/0272989X9301300107

47. Clinical Outcomes of Surgical Therapy Trial Group (2004) A comparison of laparoscopically assisted and open colectomy for colon cancer. N Engl J Med 350:2050–2059 . doi: 10.1007/s11725-012-0360-7

48. Mallett S, Royston P, Waters R, et al (2010) Reporting performance of prognostic models in cancer: a review. BMC Med 8:21 . doi: 1741-7015-8-21 [pii] 10.1186/1741-7015-8-21

49. Surveillance, Epidemiology and ERD (2017) Cancer Stat Facts: Colorectal Cancer. https://seer.cancer.gov/statfacts/html/colorect.html. Accessed 17 Mar 2018

50. Altman DG, Vergouwe Y, Royston P, Moons KGM (2009) Prognosis and prognostic research: validating a prognostic model. BMJ 338:b605 . doi: 10.1136/bmj.b605

51. Gazelle GS, Hunink MG, Kuntz KM, et al (2003) Cost-effectiveness of hepatic metastasectomy in patients with metastatic colorectal carcinoma: a state-transition Monte Carlo decision analysis. Ann Surg 237:544–555 . doi: 10.1097/01.SLA.0000059989.55280.33

52. Borie F, Combescure C, Daures JP, et al (2004) Cost-effectiveness of two follow-up strategies for curative resection of colorectal cancer: comparative study using a Markov model. World J Surg 28:563–569 . doi: 10.1007/s00268-004-7256-0

53. Erenay FS, Alagoz O, Banerjee R, et al (2016) Cost-effectiveness of alternative colonoscopy surveillance strategies to mitigate metachronous colorectal cancer incidence. Cancer 122:2560–2570 . doi: 10.1002/cncr.30091

54. Renehan AG, O'Dwyer ST, Whynes DK (2004) Cost effectiveness analysis of intensive versus conventional follow up after curative resection for colorectal cancer. BMJ 328:81 . doi: 10.1136/bmj.328.7431.81

55. Tan E, Gouvas N, Nicholls RJ, et al (2009) Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. Surg Oncol 18:15–24 . doi: S0960-7404(08)00049-2 [pii]10.1016/j.suronc.2008.05.008

56. Valk PE, Abella-Columna E, Haseman MK, et al (1999) Whole-body PET imaging with [18F]fluorodeoxyglucose in management of recurrent colorectal cancer. Arch Surg 134:503

57. Floriani I, Torri V, Rulli E, et al (2010) Performance of imaging modalities in diagnosis of liver metastases from colorectal cancer: A systematic review and meta-analysis. J Magn Reson Imaging 31:19–31 . doi: 10.1002/jmri.22010

58. Young PE, Womeldorph CM, Johnson EK, et al (2014) Early detection of colorectal cancer recurrence in patients undergoing surgery with curative intent: Current status and challenges. J Cancer 5:262–271. doi: 10.7150/jca.7988

59. Horton K, Ross A, Fishman E (2000) Spiral CT of Colon Cancer: Imaging Features and Role in Management. In: Radiographics. http://pubs.rsna.org/doi/pdf/10.1148/radiographics.20.2.g00mc14419. Accessed 10 Jan 2016

60. Frazier AL, Colditz GA, Fuchs CS, Kuntz KM (2000) Cost-effectiveness of screening for colorectal cancer in the general population. JAMA 284:1954–1961 . doi: 10.1001/jama.284.15.1954

61. Van Rijn, J. C., Reitsma, J. B., Stoker, J., Bossuyt, P. M., Van Deventer, S. J., & Dekker, E. (2006) Polyp miss rate determined by tandem colonoscopy: A systematic review. American Journal of Gastroenterology. 101(2):343-50.. https://doi.org/10.1111/j.1572-0241.2006.00390.x

62. Potter KC, Husband JE, Houghton SL, et al (2009) Diagnostic accuracy of serial ct/magnetic resonance imaging review vs. positron emission tomography/ct in colorectal cancer patients with suspected and known recurrence. Dis Colon Rectum 52:253–259 . doi: 10.1007/DCR.0b013e31819d11e6

63. Bhattacharya RN, Waymire EC (2009) Stochastic Processes with Applications. SIAM, Philadephia

64. Perez-Ocon R, Ruiz-Castro JE, Gamiz-Perez ML (2001) Non-homogeneous Markov models in the analysis of survival after breast cancer. Appl Stat 50:111–124

65. Castelli C, Combescure C, Foucher Y, Daures J-P (2007) Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. Stat Med 26:5557–5571

66. Jackson CH, Sharples LD (2002) Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. Stat Med 21:113–128

67. Dennis J, Schnabel RB (1996) Numerical Methods for Unconstrained Optimization and Nonlinear Equations. SIAM, Philadephia