Supplementary Materials

Accompanying the manuscript:

Estimating population-based recurrence rates of colorectal cancer over time in the United States

Natalia Kunst, Fernando Alarid-Escudero, Eline Aas, Veerle M.H. Coupé, Deborah Schrag, Karen M. Kuntz

Table of Contents

Supplementary Methods 2
Supplementary Results 12
Supplementary References 27

Supplementary Methods

1. SEER*Stats Code for Data Extraction

Table S1. Code used in SEER*Stats to extract data with a consistent cancer staging from 1975 through 2003 (S1,2)

CRC stage II diagnosed in 1975-1984 with tumor location in colon
(({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'
AND (({Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 0
AND (({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 4,6-9, '&'
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 0, '-', 'Blank(s)')
OR {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 1-9, '&'))
OR ({Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 0
AND {Stage - LRD (Summary and Historic).SEER historic stage A} = 'Localized'
AND (({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 4,6-9, '&'
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 0, '-', 'Blank(s)')
OR {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 1-9, '&'
OR ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 5
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = '-', 'Blank(s)')))
OR ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = '-', 'Blank(s)'
{Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.2-Digit NS EOD part 1 (1973-1982)} = 'Regional, direct extension only'
AND {Stage - LRD (Summary and Historic).SEER historic stage A} = 'Regional')))
OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984'

AND {Extent of Disease - Historic.EOD 4 - extent (1983-1987)} = 4-7
AND {Extent of Disease - Historic.EOD 4 - nodes (1983-1987)} = 0))
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Cecum', 'Ascending Colon', 'Hepatic Flexure', 'Transverse Colon', 'Splenic Flexure', 'Descending Colon', 'Sigmoid Colon', 'Large Intestine, NOS'

CRC stage II diagnosed in 1975-1984 with tumor location in rectum

(({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'
AND (({Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 0
AND (({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 4,6-9, '&'
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 0, '-', 'Blank(s)')
OR {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 1-9, '&'))
OR ({Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 0
AND {Stage - LRD (Summary and Historic).SEER historic stage A} = 'Localized'
AND (({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 4,6-9, '&'
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 0, '-', 'Blank(s)')
OR {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 1-9, '&'
OR ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 5
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = '-', 'Blank(s)')))
OR ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.2-Digit NS EOD part 1 (1973-1982)} = 'Regional, direct extension only'
AND {Stage - LRD (Summary and Historic).SEER historic stage A} = 'Regional')))
OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984'
AND {Extent of Disease - Historic.EOD 4 - extent (1983-1987)} = 4-7

AND {Extent of Disease - Historic.EOD 4 - nodes (1983-1987)} = 0))
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Rectum and Rectosigmoid Junction'

CRC stage III diagnosed in 1975-1984 with tumor location in colon

(({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'
AND ((({Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = 1
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 0)
AND ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 1-9, '&', '-', 'Blank(s)'
OR ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 2-9, '&')))
OR ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.2-Digit NS EOD part 1 (1973-1982)} = 'Regional, nodes only'
AND {Stage - LRD (Summary and Historic).SEER historic stage A} = 'Regional')))
OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984'
AND {Extent of Disease - Historic.EOD 4 - extent (1983-1987)} = 0-7
AND {Extent of Disease - Historic.EOD 4 - nodes (1983-1987)} = 1,8))
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Cecum', 'Ascending Colon', 'Hepatic Flexure', 'Transverse Colon', 'Splenic Flexure', 'Descending Colon', 'Sigmoid Colon', 'Large Intestine, NOS'

CRC stage III diagnosed in 1975-1984 with tumor location in rectum

(({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'
AND ((({Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = 1
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 0
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 0)
AND ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 1-9, '&', '-', 'Blank(s)'
OR ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = 0

AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = 2-9, '&')))
OR ({Extent of Disease - Historic.Expanded EOD(5) - CP57 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(7) - CP59 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(10) - CP62 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 'Blank(s)'
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = '-', 'Blank(s)'
AND {Extent of Disease - Historic.2-Digit NS EOD part 1 (1973-1982)} = 'Regional, nodes only'
AND {Stage - LRD (Summary and Historic).SEER historic stage A} = 'Regional')))
OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984'
AND {Extent of Disease - Historic.EOD 4 - extent (1983-1987)} = 0-7
AND {Extent of Disease - Historic.EOD 4 - nodes (1983-1987)} = 1,8))
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Rectum and Rectosigmoid Junction'

CRC stage IV diagnosed in 1975-1984 with tumor location in colon

(({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'
AND (({Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 1
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 0, '-', 'Blank(s)')
OR {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 1-9, '&'
OR {Stage - LRD (Summary and Historic).SEER historic stage A} = 'Distant'))
OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984'
AND {Extent of Disease - Historic.EOD 4 - extent (1983-1987)} = 8
OR {Extent of Disease - Historic.EOD 4 - nodes (1983-1987)} = 7))
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Cecum', 'Ascending Colon', 'Hepatic Flexure', 'Transverse Colon', 'Splenic Flexure', 'Descending Colon', 'Sigmoid Colon', 'Large Intestine, NOS'

CRC stage IV diagnosed in 1975-1984 with tumor location in rectum

(({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'
AND (({Extent of Disease - Historic.Expanded EOD(12) - CP64 (1973-1982)} = 1
AND {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 0, '-', 'Blank(s)')
OR {Extent of Disease - Historic.Expanded EOD(13) - CP65 (1973-1982)} = 1-9, '&'
OR {Stage - LRD (Summary and Historic).SEER historic stage A} = 'Distant'))

OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984'
AND {Extent of Disease - Historic.EOD 4 - extent (1983-1987)} = 8
OR {Extent of Disease - Historic.EOD 4 - nodes (1983-1987)} = 7))
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Rectum and Rectosigmoid Junction'

CRC stage II diagnosed in 1994-2003 with tumor location in colon

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1994', '1995', '1996', '1997', '1998', '1999', '2000', '2001', '2002', '2003'
AND {Stage - AJCC.AJCC stage 3rd edition (1988-2003)} = 'Stage II')
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Cecum', 'Ascending Colon', 'Hepatic Flexure', 'Transverse Colon', 'Splenic Flexure', 'Descending Colon', 'Sigmoid Colon', 'Large Intestine, NOS'

CRC stage II diagnosed in 1994-2003 with tumor location in rectum

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1994', '1995', '1996', '1997', '1998', '1999', '2000', '2001', '2002', '2003'
AND {Stage - AJCC.AJCC stage 3rd edition (1988-2003)} = 'Stage II')
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Rectum and Rectosigmoid Junction'

CRC stage III diagnosed in 1994-2003 with tumor location in colon

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1994', '1995', '1996', '1997', '1998', '1999', '2000', '2001', '2002', '2003'
AND {Stage - AJCC.AJCC stage 3rd edition (1988-2003)} = 'Stage III')
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Cecum', 'Ascending Colon', 'Hepatic Flexure', 'Transverse Colon', 'Splenic Flexure', 'Descending Colon', 'Sigmoid Colon', 'Large Intestine, NOS'

CRC stage III diagnosed in 1994-2003 with tumor location in rectum

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1994', '1995', '1996', '1997', '1998', '1999', '2000', '2001', '2002', '2003'
AND {Stage - AJCC.AJCC stage 3rd edition (1988-2003)} = 'Stage III')
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Rectum and Rectosigmoid Junction'

CRC stage IV diagnosed in 1994-2003 with tumor location in colon

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1994', '1995', '1996', '1997', '1998', '1999', '2000', '2001', '2002', '2003'
AND {Stage - AJCC.AJCC stage 3rd edition (1988-2003)} = 'Stage IV')
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Cecum', 'Ascending Colon', 'Hepatic Flexure', 'Transverse Colon', 'Splenic Flexure', 'Descending Colon', 'Sigmoid Colon', 'Large Intestine, NOS'

CRC stage IV diagnosed in 1994-2003 with tumor location in rectum

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1994', '1995', '1996', '1997', '1998', '1999', '2000', '2001', '2002', '2003'
AND {Stage - AJCC.AJCC stage 3rd edition (1988-2003)} = 'Stage IV')
AND {Site and Morphology.Site recode ICD-O-3/WHO 2008} = 'Rectum and Rectosigmoid Junction'

2. Histologic Types of the Tumor

Table S2. Histologic types of the tumor based on ICD-03 coding schema included in the analysis (S1)

SEER*Stats: ICD-0-3 Hist/behav
8000/3: Neoplasm, malignant
8001/3: Tumor cells, malignant
8010/2: Carcinoma in situ, NOS
8010/3: Carcinoma, NOS
8020/3: Carcinoma, undifferentiated, NOS
8021/3: Carcinoma, anaplastic, NOS
8022/3: Pleomorphic carcinoma
8140/3: Adenocarcinoma, NOS
8141/3: Scirrhous adenocarcinoma
8144/3: Adenocarcinoma, intestinal type
8145/3: Carcinoma, diffuse type
8210/3: Adenocarcinoma in adenomatous polyp
8211/3: Tubular adenocarcinoma
8213/3: Serrated adenocarcinoma
8220/3: Adenocarcinoma in adenomatous polyposis coli
8221/3: Adenocarcinoma in multiple adenomatous polyps
8230/3: Solid carcinoma, NOS
8255/3: Adenocarcinoma with mixed subtypes
8261/3: Adenocarcinoma in villous adenoma
8262/3: Villous adenocarcinoma
8263/3: Adenocarcinoma in tubulovillous adenoma
8480/3: Mucinous adenocarcinoma
8481/3: Mucin-producing adenocarcinoma
8490/3: Signet ring cell carcinoma
8560/3: Adenosquamous carcinoma
8570/3: Adenocarcinoma with squamous metaplasia
8574/3: Adenocarcinoma with neuroendocrine differentiation

3. Supplementary Method Description

The majority of methods to analyze time-to-event data require information about the event times. In the present study, we aimed to estimate recurrence rates of colorectal cancer (CRC) from population-based data from the Surveillance, Epidemiology, and End Results (SEER) Program (S3), which do not contain any information about recurrent disease. For this purpose, we performed continuous-time multistate survival modeling that described the process in which patients moved between three health states (no evidence of disease [NED], symptomatic recurrence [Recur], and death) in continuous time. We applied the "msm" R package that allows to fit a continuous-time multistate survival model (MSM) to data that include observations of the process at arbitrary times (i.e., unobserved exact times for when the patients transition between states) (S4). The MSMs fitted with this R package are defined by intensities (i.e., rates). The "msm" R package determines the intensities with both the time that individuals spent in the current state and the next state probability. In the present study, our MSMs estimated transition intensities between the model's health states using a maximum likelihood estimation approach (S4). When information about the observed events is available in the data (i.e., the exact transition times are known), the likelihood is estimated with a transition intensity matrix Λ. If the observations of the Markov process are at arbitrary times (i.e., the exact transition times are unknown), the likelihood is estimated in terms of the transition probability matrix $P(u) = exp(u\Lambda)$, which is a matrix exponential of the scaled transition intensity matrix (see Cox and Miller (S5) for more detailed explanation) (S4). When considering time-homogenous process, the (i, j) entry of the transition probability matrix $P(q, u + q)$ represents the probability of state j at a future time $u + q$, conditioning on i being the state at time q (S4). Given the constant transition intensity matrix Λ in the time interval $(q, u + q)$ in time-homogenous process, the transition probability matrix equals to $P(q, u + q) = P(u)$. For more detailed description about the likelihood estimation and methods used in the "msm" R package, please see Jackson (S6), Jackson (S4), Kalbfleisch and Lawless (S7), Kay (S8), and Gentleman, et al. (S9).

In our model with three health states, there are three possible transitions: transitions from NED to Recur, transitions from NED to death due to other causes, and transitions from Recur to death. We informed the transitions λ_{12}^{ct}, where c denotes cohort of patients diagnosed either between 1975-1984 or 1994-2003 and t denotes time since diagnosis for piecewise-constant rates (6 months-5 years and 5-10 years post-diagnosis), with observations of the process at arbitrary times (unobserved events in the SEER data). The transitions λ_{23}^{ct} were informed using exact times observed in the SEER data on patients with *de novo* stage IV CRC, assuming that those patients and patients in the Recur state would have a similar risk of death based on the previous literature (S10). The transitions λ_{13}^{ct} were informed using estimates on the expected survival for each of the subpopulations considered in our analysis. These estimates were

obtained from SEER*Stat from the National Cancer Institute (S1). The SEER Program calculates these estimates by constructing a non-CRC cohort matched to the CRC cohort by patients' age, sex and race (S1). Using these data, we estimated annual all-cause mortality rates for each of the subpopulations analyzed in our study. These values were used as the initial values (see more detailed description below) and were fixed during the optimization, assuming a constant annual all-cause mortality rate. This assumption was necessary to fit the MSMs to the available data. For censored patients, that is, patients that were alive at the study cutoff, patients' final observation was unknown (i.e., NED or Recur). Thus, we used so-called "censored" states. This terminology is used in the "msm" R package. The censored states in this package are used in situations when we only know that patients' final observation is either NED or Recur and we need to assign patients to one of these states. The "msm" R package uses a modification of the method provided by Titman and Sharples (S11) to estimate the likelihood in presence of censored states. Detailed information about the calculation of the likelihood for such censored observations is provided in Jackson (S4).

In the present study, we created 18 distinct MSMs for patient samples defined by age (20-49, 50-64, and 65-79 years), stage and location (stage II colon, stage II rectum, stage III), and diagnosis periods (i.e., 1975-1984 and 1994-2003). Each of the MSMs had a transition matrix of

$$\Lambda^{ct}=\begin{pmatrix} -(\lambda_{12}^{ct}+\lambda_{13}^{ct}) & \lambda_{12}^{ct} & \lambda_{13}^{ct} \\ 0 & -\lambda_{23}^{ct} & \lambda_{23}^{ct} \\ 0 & 0 & 0 \end{pmatrix},$$

where *c* denotes cohort of patients diagnosed either between 1975 and 1984 or 1994 and 2003, and *t* denotes time since diagnosis for piecewise-constant rates, that is, 6 months to 5 years or 5 years to 10 years post-diagnosis. The process of fitting the model is performed by finding the three unknown transition intensities (λ_{12}^{ct}, λ_{13}^{ct}, and λ_{23}^{ct}) that maximize the likelihood.

The "msm" R package allows to fit time-inhomogeneous MSMs where the intensities change with time. To estimate *P(u)* for time-inhomogeneous process, time-dependent covariates are assumed piecewise-constant with exponential distribution. Because time-dependency is incorporated in the model in form of covariates, and we were interested in obtaining estimates of the CRC recurrence rates for each of the specified time periods rather than an average of the piecewise-constant intensities across 10 years, it was not feasible to fit the other covariates, including sex and race, in these time-inhomogeneous MSMs. However, we also created time-homogeneous MSMs fitted with sex, race, and tumor location covariates. The results of these MSMs estimating 10-year constant recurrence rates can be obtained upon request. The piecewise-constant MSMs, which were created to better reflect clinical knowledge, allowed transition rates to change over the 10-post-diagnosis years. Based on goodness-of-fit tests and visual inspection of

models, we estimated piecewise-constant recurrence rates for 6 months-5 years and 5-10 years post-diagnosis. These rates were estimated separately for patient with CRC stage II or III in all age groups, and tumor location in CRC stage II, in two diagnosis periods (i.e., 1975-1984 and 1994-2003), and provided more detailed information about patients' disease patterns.

As previously mentioned, the "msm" R package applies a maximum likelihood estimation approach to estimate intensities in the MSMs. To maximize the likelihood using numerical methods, initial values for all non-zero λ^{ct}included in the transition matrix Λ^{ct} should be specified to start the search for the maximum (S4,6). As recommended, we ran the MSMs using a number of different initial values (S4). It is further recommended to specify reasonable initial values for more complex or unstable models (S4). Thus, we used a number of sources to estimate the initial values for our transition intensities λ_{12}^{ct}, λ_{13}^{ct}, and λ_{23}^{ct}, and the initial values for log-linear effect of our time-dependent covariates, that is, different initial values depending on the time since diagnosis. Detailed explanation is provided in Jackson (S4) and Jackson (S6). We informed the transition intensity λ_{12}^{ct} (i.e., transition rate from NED to Recur) with the recurrence estimates provided by Sargent, et al. (S12). Although this study provided trial-based recurrence estimates, to our knowledge, this was the most appropriate source as it provided detailed recurrence estimates for every six post-diagnosis months over 8 post-diagnosis years. We assumed a constant estimate of the recurrence rates after 8 years post-diagnosis (i.e., between 8-10 post-diagnosis years). Using these recurrence estimates, we estimated average recurrence rates for 6 months-5years post-diagnosis and 5-10 years post-diagnosis, and used these estimates as the initial values in our MSMs. Next, we informed the transition intensity λ_{23}^{ct} (i.e., transition rates from Recur to Death) with SEER data. More specifically, for each of the subpopulations analyzed in our study (i.e., patients stratified by diagnosis period, age, disease stage, and tumor location [CRC stage II]), we obtained estimates for the observed survival of patients with CRC *de novo* stage IV from SEER*Stats. Using these estimates, we first estimated mortality rates for each post-diagnosis month. Thereafter, we estimated average mortality rates for 6 months-5 years post-diagnosis and 5-10 years post-diagnosis, and used these estimates as the initial values in our MSMs. Finally, we informed the transition intensity λ_{13}^{ct} in a similar way as the transition intensity λ_{23}^{ct} but instead of using the observed survival, we used the expected survival for each of the analyzed subpopulations with CRC stage II or III obtained from SEER*Stats (i.e., patients stratified by diagnosis period, age and cancer stage). Using these expected survival estimates, we first estimated mortality rates for each post-diagnosis month. Next, we estimated average mortality rates for 6 months-5 years post-diagnosis and 5-10 years post-diagnosis, and used these estimates as the initial values in our MSMs.

Supplementary Results

Table S3. Population-based annual recurrence rates in patient with stage II CRC (combined colon and rectum) diagnosed in 1975-1984 or 1994-2003 by post-diagnosis time

	Time since diagnosis	
	6 months to 5 years	**5 years to 10 years**
	Rate (95% CI)	**Rate (95% CI)**
Diagnosis in 1975-1984		
Age 20-49	0.065 (0.055-0.077)	0.049 (0.026-0.095)
Age 50-64	0.069 (0.064-0.075)	0.045 (0.032-0.064)
Age 65-79	0.064 (0.060-0.069)	0.076 (0.049-0.117)
Diagnosis in 1994-2003		
Age 20-49	0.038 (0.030-0.048)	0.027 (0.008-0.089)
Age 50-64	0.039 (0.033-0.046)	0.036 (0.015-0.086)
Age 65-79	0.044 (0.039-0.050)	0.042 (0.013-0.138)

Table S4. Hazard ratios of recurrence recurrences in patients with colorectal cancer by post-diagnosis time, age group, and stage for patients diagnosed in 1994-2003 vs 1975-1984

	Time since diagnosis	
	6 months to 5 years	**5 years to 10 years**
Age	**HR (95% CI)**	**HR (95% CI)**
Stage II		
Colon		
Age 20-49	0.646 (0.542-0.770)	0.546 (0.280-1.065)
Age 50-64	0.553 (0.505-0.605)	0.813 (0.476-1.388)
Age 65-79	0.702 (0.645-0.764)	0.949 (0.391-2.301)
Rectum		
Age 20-49	0.428 (0.359-0.510)	0.362 (0.207-0.633)
Age 50-64	0.595 (0.534-0.664)	0.875 (0.492-1.556)
Age 65-79	0.672 (0.607-0.744)	0.908 (0.383-2.151)
Stage III		
Age 20-49	0.673 (0.589-0.770)	0.583 (0.107-3.174)
Age 50-64	0.622 (0.583-0.663)	0.969 (0.216-4.353)
Age 65-79	0.598 (0.559-0.639)	0.787 (0.420-1.474)

Table S5. Population-based annual recurrence rates in patient aged 20-39 and 40-49 with CRC stage II colon, stage II rectum, and stage III diagnosed in 1975-1984 and 1994-2003 by time since diagnosis

	Time since diagnosis	
	6 months to 5 years	**5 years to 10 years**
	Recurrence rate (95% CI)	**Recurrence rate (95% CI)**
Diagnosis in 1975-1984		
Stage II		
Age 20-39		
Colon	0.046 (0.030-0.070)	0.042 (0.006-0.278)
Rectum	0.073 (0.043-0.123)	0.067 (0.010-0.459)
Age 40-49		
Colon	0.056 (0.045-0.068)	0.034 (0.015-0.079)
Rectum	0.116 (0.091-0.149)	0.071 (0.030-0.169)
Stage III		
Age 20-39	0.114 (0.071-0.184)	*
Age 40-49	0.152 (0.129-0.180)	0.022 (0.004-0.121)
Diagnosis in 1994-2003		
Stage II		
Age 20-39		
Colon	0.021 (0.010-0.043)	0.058 (0.011-0.325)
Rectum	0.038 (0.018-0.083)	0.106 (0.017-0.649)
Age 40-49		
Colon	0.040 (0.030-0.053)	0.017 (0.003-0.099)
Rectum	0.046 (0.033-0.066)	0.020 (0.003-0.115)
Stage III		
Age 20-39	0.086 (0.056-0.133)	0.075 (0.003-1.688)
Age 40-49	0.099 (0.083-0.119)	*

*The model failed to find the maximum of the log-likelihood (i.e., failure to converge). Additional stratification might result in an overly-complex model with insufficient data failing to identify the parameters of the model.

Figure S1. Estimates of relative survival with 95% confidence intervals among patients aged 20-39 and 40-49 years diagnosed with stage II colon, stage II rectum, or stage III colorectal cancer in 1975-1984 and 1994-2003

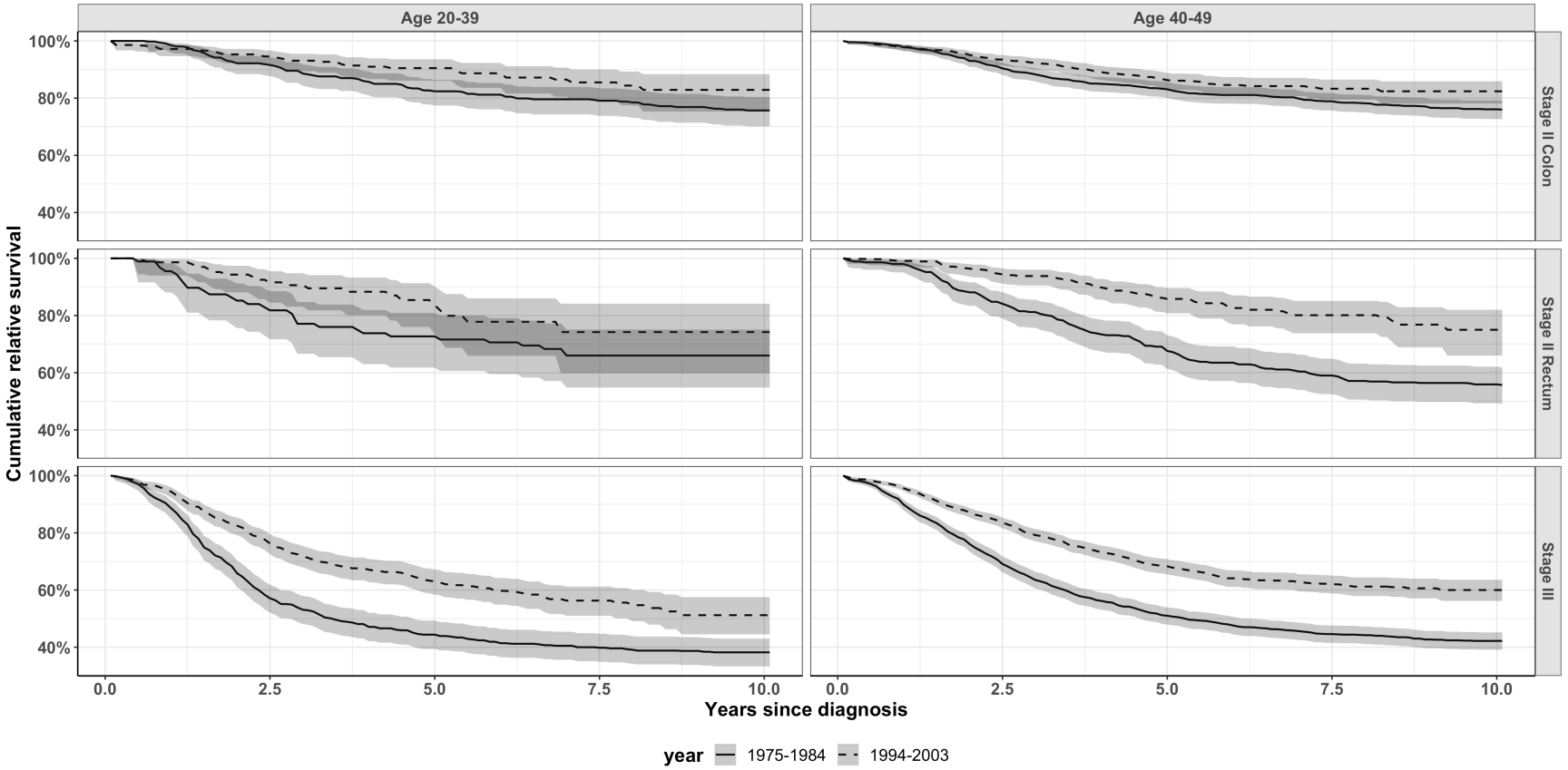


Table S6. Population-based annual recurrence rates in patient with CRC stage II colon or stage III colon diagnosed in 1975-1984 and 1994-2003 by time since diagnosis and colon cancer location[ab]

	Right colon cancer[a]		Left colon cancer[b]	
	Time since diagnosis			
	6 months to 5 years	5 years to 10 years	6 months to 5 years	5 years to 10 years
	Recurrence rate (95% CI)	Recurrence rate (95% CI)	Recurrence rate (95% CI)	Recurrence rate (95% CI)
Diagnosis in 1975-1984				
Stage II				
Age 20-49	0.044 (0.032-0.060)	*	0.062 (0.046-0.084)	0.040 (0.014-0.109)
Age 50-64	0.047 (0.040-0.056)	0.029 (0.013-0.062)	0.066 (0.057-0.075)	0.049 (0.025-0.097)
Age 65-79	0.038 (0.032-0.044)	0.022 (0.004-0.121)	0.072 (0.064-0.082)	0.064 (0.020-0.207)
Stage III				
Age 20-49	0.112 (0.079-0.159)	*	0.151 (0.114-0.199)	*
Age 50-64	0.125 (0.106-0.146)	0.016 (0.003-0.096)	0.172 (0.154-0.193)	0.026 (0.005-0.123)
Age 65-79	0.106 (0.090-0.124)	0.079 (0.036-0.177)	0.156 (0.138-0.177)	0.111 (0.046-0.264)
Diagnosis in 1994-2003				
Stage II				
Age 20-49	0.023 (0.014-0.038)	*	0.046 (0.030-0.070)	0.063 (0.011-0.377)
Age 50-64	0.022 (0.015-0.032)	0.018 (0.003-0.102)	0.043 (0.033-0.056)	0.035 (0.008-0.148)
Age 65-79	0.029 (0.024-0.037)	*	0.055 (0.045-0.067)	0.026 (0.002-0.287)
Stage III				
Age 20-49	0.075 (0.051-0.111)	*	0.109 (0.084-0.144)	*
Age 50-64	0.089 (0.072-0.110)	*	0.096 (0.080-0.115)	*

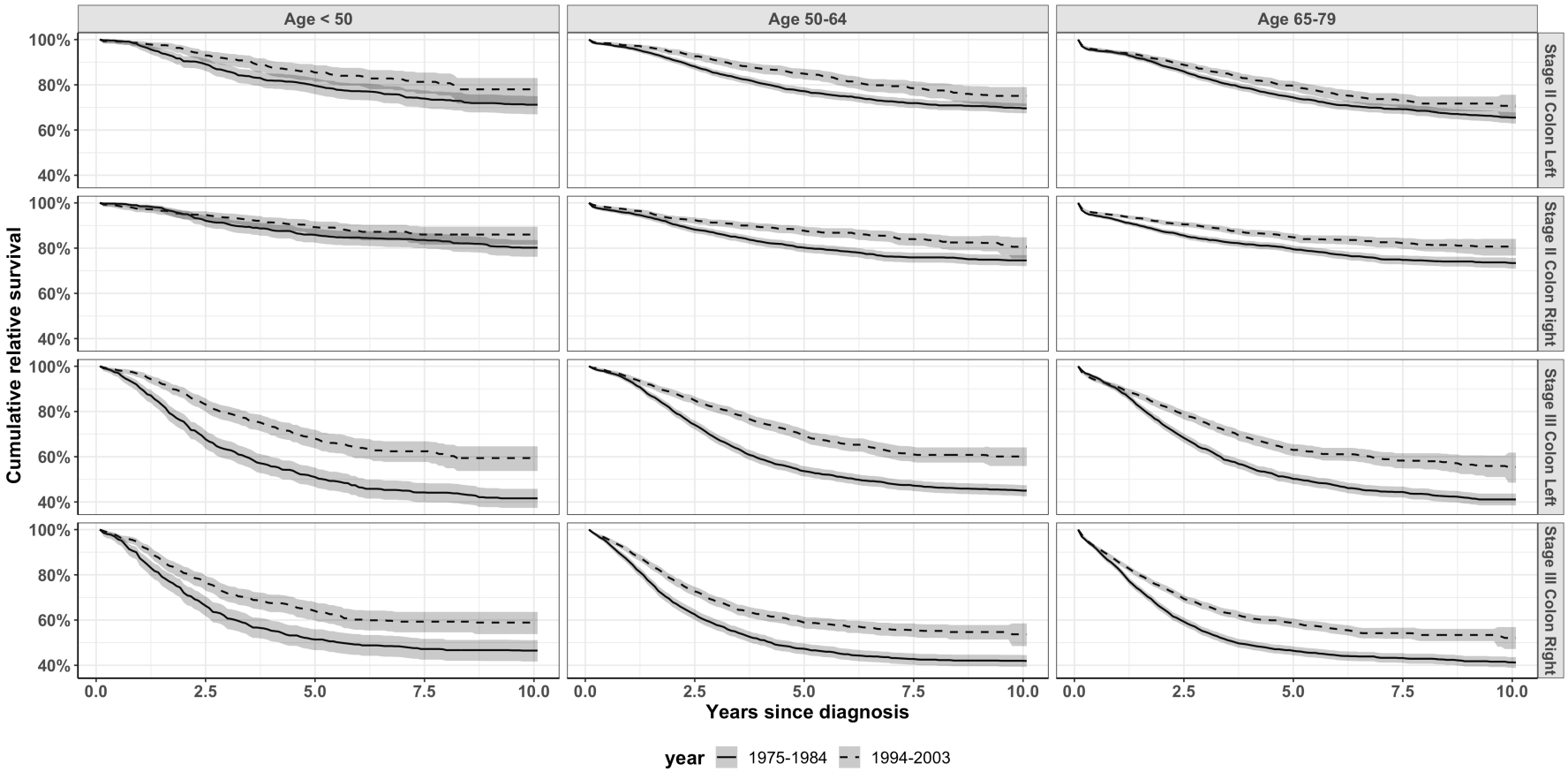
Age 65-79	0.069 (0.057-0.084)	0.054 (0.005-0.652)	0.095 (0.079-0.115)	0.051 (0.004-0.633)

[a]Right colon cancer includes cecum, ascending colon, transverse colon, and hepatic flexure.

[b]Left colon cancer includes descending colon, sigmoid colon, and splenic flexure.

*The model failed to find the maximum of the log-likelihood (i.e., failure to converge). Additional stratification might result in an overly-complex model with insufficient data failing to identify the parameters of the model.

Figure S2. Estimates of relative survival with 95% confidence intervals among patients aged 20-49, 50-64, and 65-79 with CRC stage II colon or stage III colon diagnosed in 1975-1984 and 1994-2003 by time since diagnosis and colon cancer location[ab]



[a]Right colon cancer includes cecum, ascending colon, transverse colon, and hepatic flexure.

[b]Left colon cancer includes descending colon, sigmoid colon, and splenic flexure.

Table S7. Population-based annual recurrence rates in patient with CRC stage II colon, stage II rectum, and stage III diagnosed in 1975-1984 and 1994-2003 by time since diagnosis and patient race

	White		Non-white[a]	
	Time since diagnosis			
	6 months to 5 years	5 years to 10 years	6 months to 5 years	5 years to 10 years
	Recurrence rate (95% CI)	Recurrence rate (95% CI)	Recurrence rate (95% CI)	Recurrence rate (95% CI)
Diagnosis in 1975-1984				
Stage II				
Age 20-49				
Colon	0.058 (0.048-0.071)	0.038 (0.017-0.082)	0.034 (0.018-0.063)	0.030 (0.003-0.269)
Rectum	0.112 (0.088-0.142)	0.072 (0.032-0.161)	0.071 (0.035-0.144)	0.062 (0.007-0.056)
Age 50-64				
Colon	0.058 (0.053-0.064)	0.041 (0.029-0.059)	0.074 (0.059-0.092)	0.030 (0.008-0.103)
Rectum	0.091 (0.082-0.102)	0.064 (0.045-0.092)	0.123 (0.094-0.161)	0.049 (0.014-0.175)
Age 65-79				
Colon	0.057 (0.052-0.061)	0.033 (0.016-0.066)	0.059 (0.054-0.079)	*
Rectum	0.097 (0.088-0.108)	0.056 (0.028-0.114)	0.079 (0.054-0.115)	*
Stage III				
Age 20-49	0.144 (0.120-0.172)	0.032 (0.008-0.138)	0.155 (0.113-0.212)	*
Age 50-64	0.179 (0.167-0.192)	0.014 (0.004-0.056)	0.162 (0.136-0.194)	0.029 (0.006-0.151)
Age 65-79	0.217 (0.182-0.259)	0.124 (0.017-0.887)	0.149 (0.138-0.162)	0.083 (0.050-0.136)
Diagnosis in 1994-2003				
Stage II				
Age 20-49				

Colon	0.029 (0.021-0.040)	0.025 (0.007-0.091)	0.055 (0.036-0.084)	*
Rectum	0.043 (0.029-0.063)	0.038 (0.010-0.138)	0.056 (0.032-0.099)	*
Age 50-64				
Colon	0.031 (0.026-0.038)	0.028 (0.008-0.100)	0.041 (0.029-0.057)	0.021 (0.003-0.147)
Rectum	0.055 (0.044-0.069)	0.049 (0.014-0.177)	0.063 (0.042-0.096)	0.032 (0.005-0.228)
Age 65-79				
Colon	0.040 (0.035-0.046)	0.024 (0.005-0.114)	0.038 (0.027-0.053)	*
Rectum	0.065 (0.055-0.077)	0.039 (0.008-0.186)	0.054 (0.035-0.082)	*
Stage III				
Age 20-49	0.089 (0.072-0.110)	0.021 (0.002-0.191)	0.126 (0.096-0.164)	*
Age 50-64	0.109 (0.097-0.122)	0.019 (0.003-0.133)	0.114 (0.092-0.140)	*
Age 65-79	0.088 (0.079-0.099)	0.060 (0.024-0.175)	0.122 (0.100-0.149)	*

[a]Due to a high proportion of white patients in the sample, we classified all patients into white and non-white race. The non-white race group included patients recorded as black and other (American Indian/Alaskan Native and Asian/Pacific Islander).

*The model failed to find the maximum of the log-likelihood (i.e., failure to converge). Additional stratification might result in an overly-complex model with insufficient data failing to identify the parameters of the model.

Figure S3. Estimates of relative survival with 95% confidence intervals among white patient aged 20-49, 50-64, and 65-79 with CRC stage II colon, stage II rectum, and stage III diagnosed in 1975-1984 and 1994-2003 by time since diagnosis

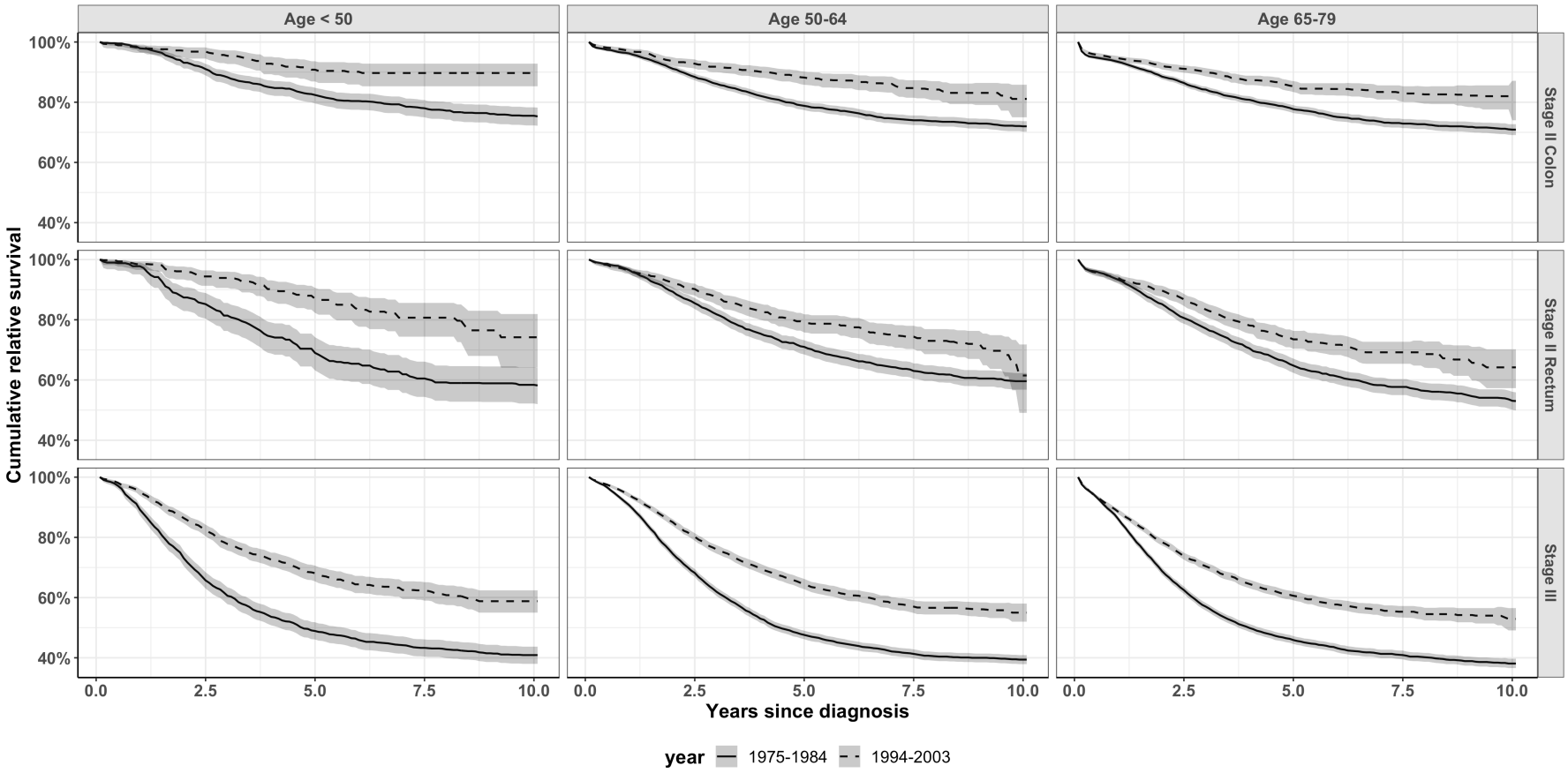


Figure S4. Estimates of relative survival with 95% confidence intervals among non-white patient aged 20-49, 50-64, and 65-79 with CRC stage II colon, stage II rectum, and stage III diagnosed in 1975-1984 and 1994-2003 by time since diagnosis

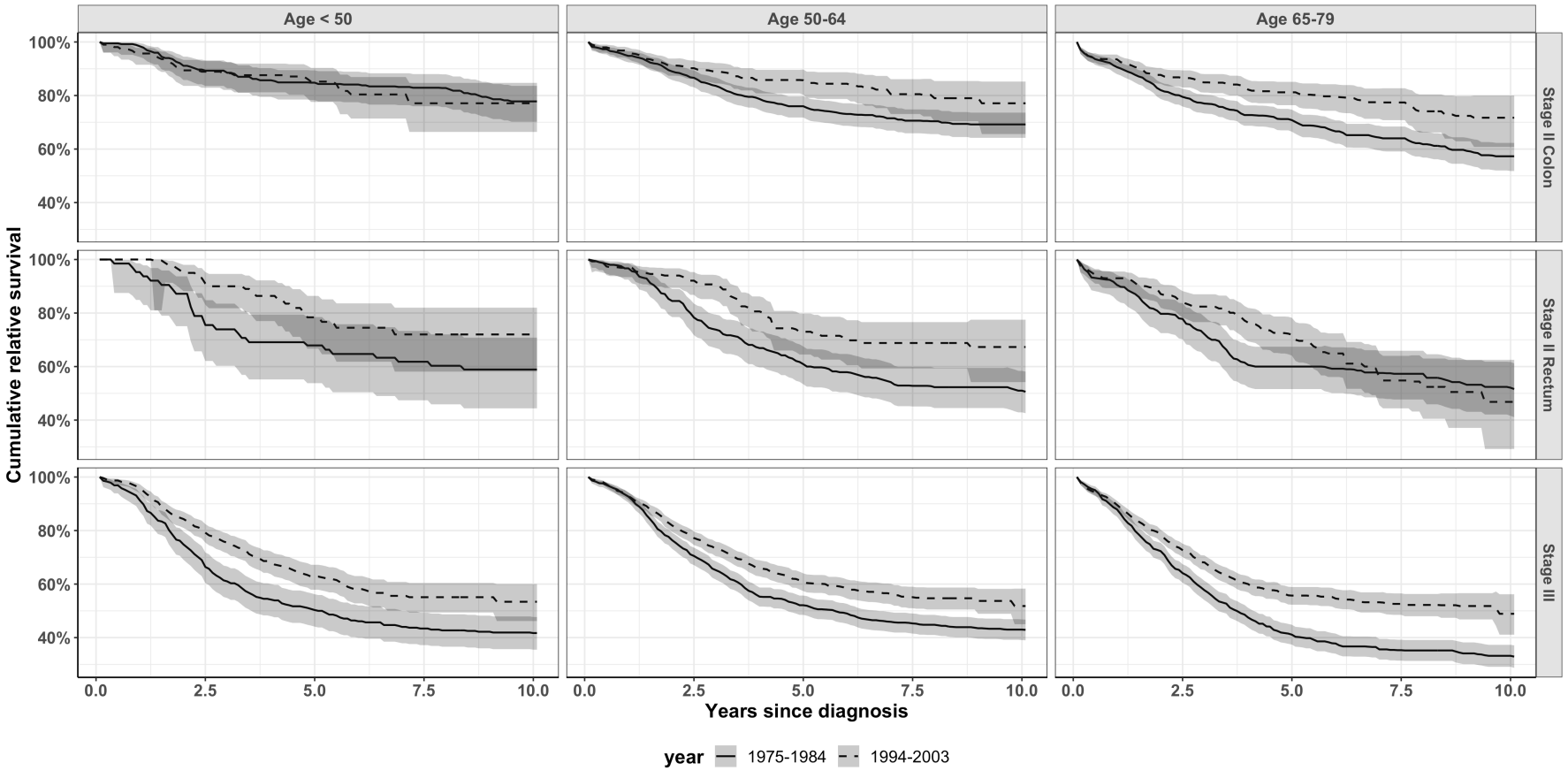


Table S8. Population-based annual recurrence rates in patient with CRC stage II colon, stage II rectum, and stage III diagnosed in 1975-1984 and 1994-2003 by time since diagnosis and patient sex

	Female		Male	
	Time since diagnosis			
	6 months to 5 years	5 years to 10 years	6 months to 5 years	5 years to 10 years
	Recurrence rate (95% CI)	Recurrence rate (95% CI)	Recurrence rate (95% CI)	Recurrence rate (95% CI)
Diagnosis in 1975-1984				
Stage II				
Age 20-49				
Colon	0.056 (0.044-0.073)	0.028 (0.010-0.077)	0.048 (0.036-0.063)	0.037 (0.005-0.295)
Rectum	0.120 (0.087-0.165)	0.035 (0.014-0.083)	0.103 (0.091-0.117)	0.070 (0.047-0.105)
Age 50-64				
Colon	0.058 (0.053-0.064)	0.041 (0.029-0.059)	0.074 (0.059-0.092)	0.030 (0.008-0.103)
Rectum	0.091 (0.082-0.102)	0.064 (0.045-0.092)	0.123 (0.094-0.161)	0.049 (0.014-0.175)
Age 65-79				
Colon	0.051 (0.045-0.059)	0.023 (0.009-0.054)	0.068 (0.061-0.076)	0.047 (0.032-069)
Rectum	0.078 (0.066-0.094)	0.035 (0.014-0.083)	0.103 (0.091-0.117)	0.070 (0.047-0.105)
Stage III				
Age 20-49	0.133 (0.107-0.165)	*	0.161 (0.129-0.201)	0.074 (0.017-0.326)
Age 50-64	0.149 (0.133-0.167)	0.037 (0.012-0.111)	0.198 (0.183-0.215)	0.010 (0.001-0.069)
Age 65-79	0.121 (0.107-0.138)	0.060 (0.023-0.154)	0.193 (0.176-0.210)	0.086 (0.044-0.170)
Diagnosis in 1994-2003				
Stage II				
Age 20-49				

Colon	0.034 (0.023-0.051)	0.022 (0.003-0.171)	0.037 (0.026-0.053)	0.021 (0.004-0.107)
Rectum	0.039 (0.022-0.069)	0.025 (0.003-0.205)	0.048 (0.032-0.070)	0.027 (0.005-0.138)
Age 50-64				
Colon	0.029 (0.022-0.038)	0.022 (0.003-0.146)	0.037 (0.030-0.046)	0.029 (0.009-0.097)
Rectum	0.053 (0.038-0.073)	0.040 (0.006-0.274)	0.058 (0.045-0.074)	0.045 (0.013-0.151)
Age 65-79				
Colon	0.026 (0.021-0.033)	0.011 (0.0002-0.675)	0.055 (0.048-0.064)	0.029 (0.001-0.608)
Rectum	0.040 (0.029-0.054)	0.016 (0.0003-1.038)	0.086 (0.072-0.103)	0.045 (0.002-0.949)
Stage III				
Age 20-49	0.114 (0.091-0.143)	*	0.087 (0.068-0.112)	0.014 (0.0006-0.326)
Age 50-64	0.096 (0.082-0.112)	0.027 (0.004-0.177)	0.122 (0.107-0.138)	0.011 (0.0004-0.371)
Age 65-79	0.069 (0.058-0.083)	0.040 (0.007-0.215)	0.118 (0.105-0.134)	0.094 (0.028-0.308)

*The model failed to find the maximum of the log-likelihood (i.e., failure to converge). Additional stratification might result in an overly-complex model with insufficient data failing to identify the parameters of the model.

Figure S5. Estimates of relative survival with 95% confidence intervals among female patients aged 20-49, 50-64, and 65-79 with CRC stage II colon, stage II rectum, and stage III diagnosed in 1975-1984 and 1994-2003 by time since diagnosis

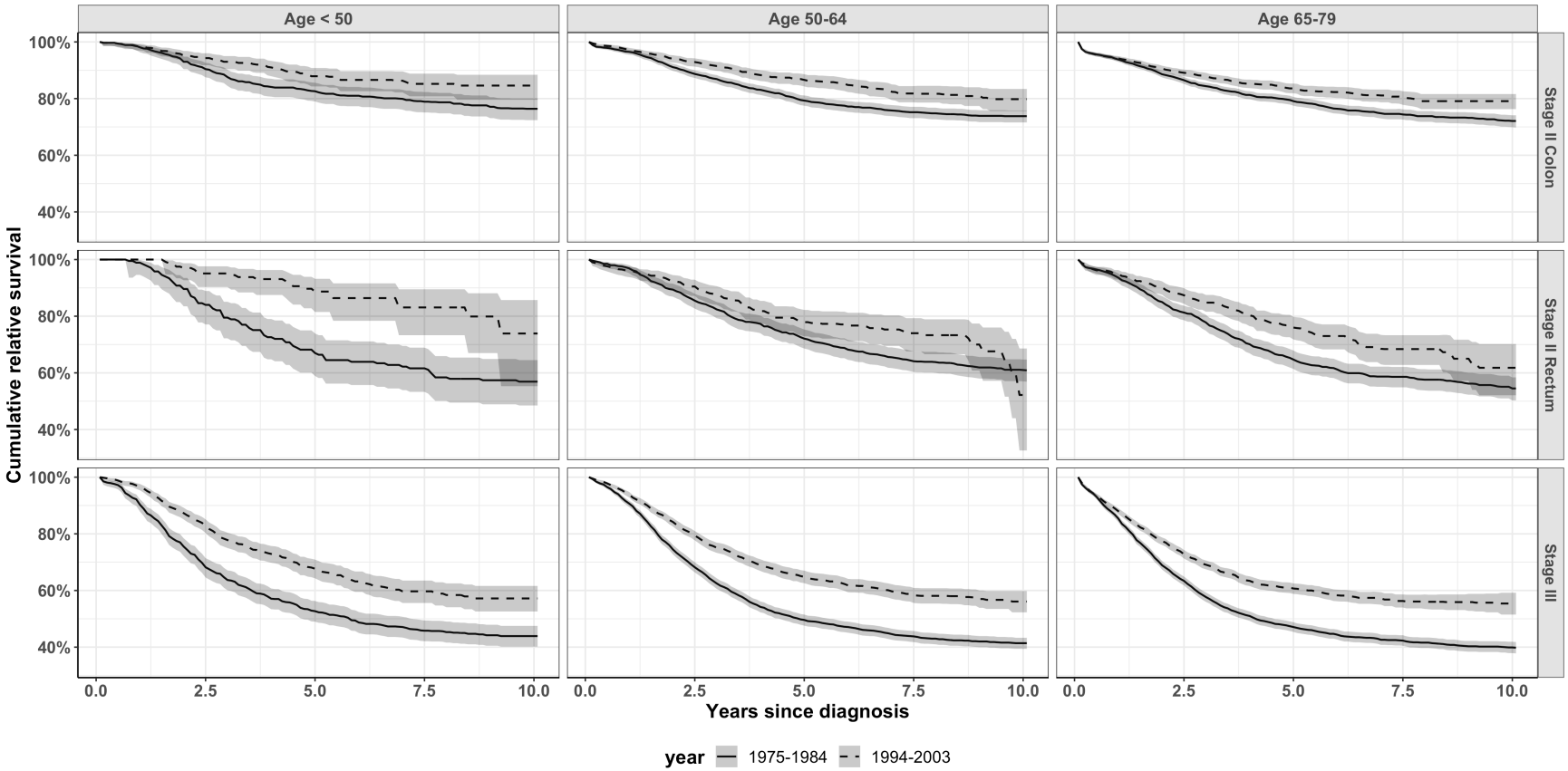
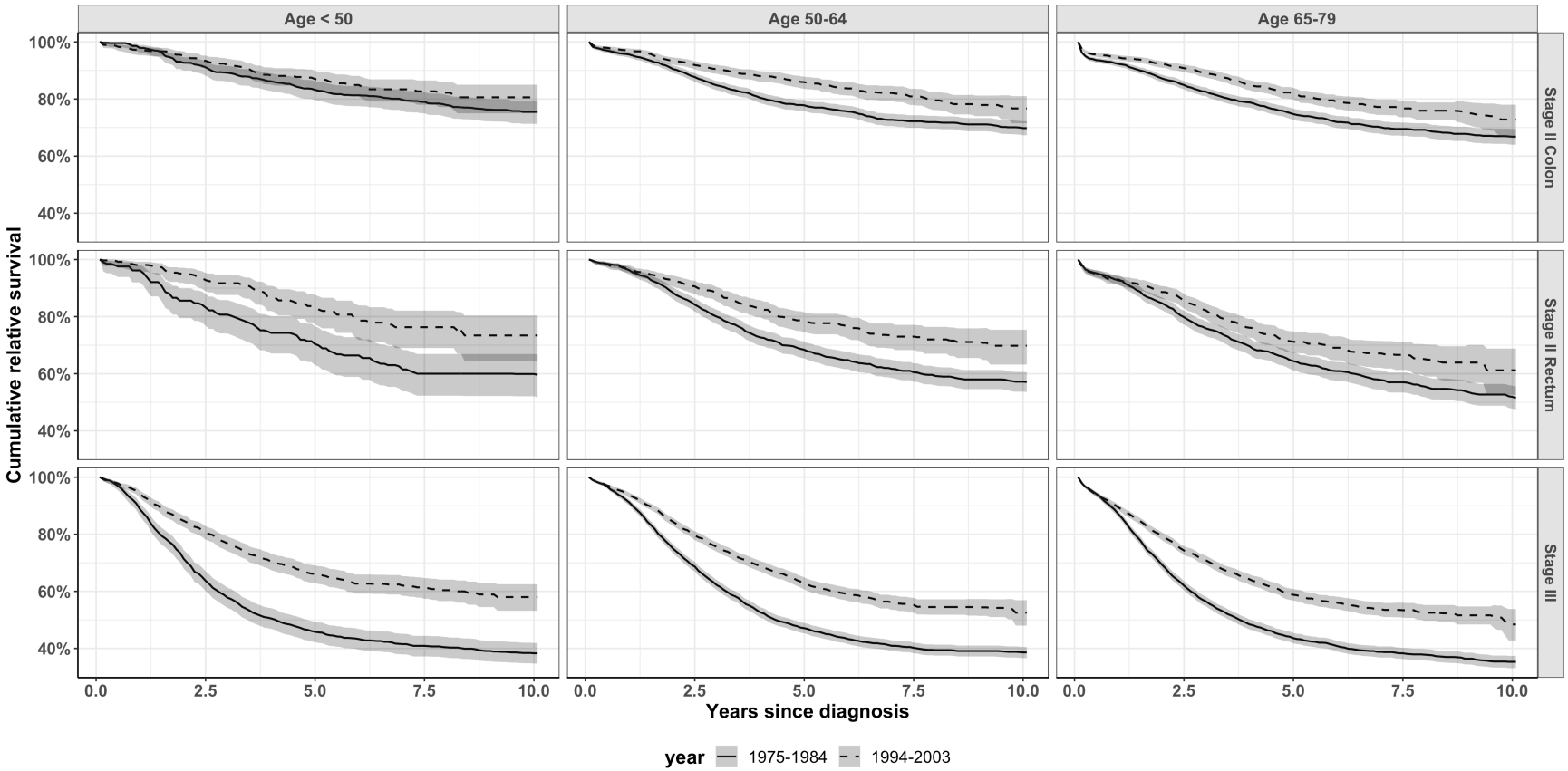


Figure S6. Estimates of relative survival with 95% confidence intervals among male patients aged 20-49, 50-64, and 65-79 with CRC stage II colon, stage II rectum, and stage III diagnosed in 1975-1984 and 1994-2003 by time since diagnosis



Supplementary References

S1. National Cancer Institute. SEER*Stat Software. https://seer.cancer.gov/seerstat/. Accessed April 7, 2019.

S2. Schrag D. AJCC Staging: Staging Colon Cancer Patients Using the TNM System From 1975 to the Present. http://www.mskcc.org/mskcc/html/84761.cfm. Accessed August 6, 2019.

S3. National Cancer Institute. Cancer Incidence—Surveillance, Epidemiology, and End Results (SEER) Registries Research Data. https://seer.cancer.gov/data-software/documentation/seerstat/nov2018/.

S4. Jackson C. Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software.* 2011;38(8):1-28.

S5. Cox DR, Miller HD. *Theory of Stochastic Processes.* London: Chapman and Hall; 1965.

S6. Jackson C. Package 'msm': Multi-State Markov and Hidden Markov Models in Continuous Time. 2019; https://cran.rstudio.com/web/packages/msm/msm.pdf. Accessed November 15, 2019.

S7. Kalbfleisch JD, Lawless JF. The Analysis of Panel Data Under a Markov Assumption. *Journal of the American Statistical Association.* 1985;80(392):863-871.

S8. Kay R. A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics.* 1986;42(4):855-865.

S9. Gentleman RC, Lawless JF, Lindsey JC, Yan P. Multi-state Markov models for analysing incomplete disease history data with illustrations for hiv disease. *Stat Med.* 1994;13(8):805-821.

S10. Hassett MJ, Uno H, Cronin AM, Carroll NM, Hornbrook MC, Ritzwoller DP. Comparing Survival After Recurrent vs De Novo Stage IV Advanced Breast, Lung, and Colorectal Cancer. *JNCI cancer spectrum.* 2018;2(2):pky024.

S11. Titman AC, Sharples LD. A general goodness-of-fit test for Markov and hidden Markov models. *Stat Med.* 2008;27(12):2177-2195.

S12. Sargent DJ, Patiyil S, Yothers G, et al. End points for colon cancer adjuvant trials: observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. *J Clin Oncol.* 2007;25(29):4569-4574.