Published in final edited form as:

J Am Coll Surg. 2016 March ; 222(3): 232–244. doi:10.1016/j.jamcollsurg.2015.12.019.

Predicting Individualized Postoperative Survival for Stage II/III Colon Cancer Using a Mobile Application Derived from the National Cancer Data Base

Emmanuel Gabriel, MD, PhD, Kristopher Attwood, PhD, Pragatheeshwar Thirunavukarasu, MD, Eisar Al-Sukhni, MD, Patrick Boland, MD, and Steven Nurkin, MD, FACS

Departments of Surgical Oncology (Gabriel, Thirunavukarasu, Al-Sukhni, Nurkin), Biostatistics (Attwood), and Medical Oncology (Boland), Roswell Park Cancer Institute, Buffalo, NY

Abstract

BACKGROUND—Prediction calculators estimate postoperative survival and assist the decision-making process for adjuvant treatment. The objective of this study was to create a postoperative overall survival (OS) calculator for patients with stage II/III colon cancer. Factors that influence OS, including comorbidity and postoperative variables, were included.

STUDY DESIGN—The National Cancer Data Base was queried for patients with stage II/III colon cancer, diagnosed between 2004 and 2006, who had surgical resection. Patients were randomly divided to a testing (n_t) cohort comprising 80% of the dataset and a validation (n_v) cohort comprising 20%. Multivariable Cox proportional hazards regression of n_t was performed to identify factors associated with 5-year OS. These were used to build a prediction model. The performance was assessed using the n_v cohort and translated into mobile software.

RESULTS—A total of 129,040 patients had surgery. After exclusion of patients with carcinoma in situ, non-adenocarcinoma histology, more than 1 malignancy, stage I or IV disease, or missing data, 34,176 patients were used in the development of the calculator. Independent predictors of OS included patient-specific characteristics, pathologic factors, and treatment options, including type of surgery and adjuvant therapy. Length of postoperative stay and unplanned readmission rates were also incorporated as surrogates for postoperative complications (1-day increase in postoperative stay, hazard ratio [HR] 1.019, 95% CI 1.018 to 1.021, $p < 0.001$; unplanned readmission vs no readmission HR 1.35, 95% CI 1.25 to 1.45, $p < 0.001$). Predicted and actual 5-year OS rates were compared in the n_v cohort with 5-year area under the curve of 0.77.

Correspondence address: Steven Nurkin, MD, FACS, Department of Surgical Oncology, Roswell Park Cancer Institute, Elm and Carlton Sts, Buffalo, NY 14263. steven.nurkin@roswellpark.org.

Author Contributions

Study conception and design: Gabriel, Thirunavukarasu, Boland, Nurkin

Acquisition of data: Gabriel, Thirunavukarasu, Al-Sukhni Analysis and interpretation of data: Gabriel, Attwood Drafting of manuscript: Gabriel, Attwood, Al-Sukhni Critical revision: Attwood, Thirunavukarasu, Boland, Nurkin

CME questions for this article available at <http://jacscme.facs.org>

Disclaimer: The American College of Surgeons Commission on Cancer provided the Participant User File from the National Cancer Data Base, but has not reviewed or validated the results or conclusions of this study.

Disclosure Information: Authors have nothing to disclose. Timothy J Eberlein, Editor-in-Chief, has nothing to disclose.

CONCLUSIONS—An individualized, postoperative OS calculator application was developed for patients with stage II/III colon cancer. This prediction model uses nationwide data, culminating in a highly comprehensive, clinically useful tool.

Overall survival (OS) for patients with colon cancer has been shown to be significantly associated with stage of disease.^{1,2} In addition to cancer stage, many factors specific to colon cancer have been shown to affect long-term OS. These factors include tumor size, histologic grade, presence of lymphovascular invasion (LVI) or perineural invasion (PNI), adequacy of oncologic resection including margin status and the number of lymph nodes acquired in the specimen, and the type of adjuvant treatment received.^{3–8}

Several groups have incorporated many of these variables into mathematical models to predict the individualized, long-term survival for a given patient with colon cancer.^{9–12} For some of these calculators, the patient cohorts used to generate the prediction models were derived from individual institution records, datasets pooled from multiple adjuvant trials (such as the Adjuvant Colon Cancer End Points or ACCENT dataset) or from large nationwide databases (namely, the Adjuvant! Online calculator derived from the Surveillance, Epidemiology, and End Results [SEER] database).^{11–13} There are other factors that contribute to OS, including patient age and other demographics including race, patient-specific comorbid conditions, elevated levels of CEA, and complications from surgery or prolonged postoperative stay.^{14–16} Unlike pathologic features, some of these other items have not been incorporated into current survival calculators.

The National Cancer Data Base (NCDB) is a large database jointly maintained by the Commission on Cancer (CoC) of the American College of Surgeons (ACS) and the American Cancer Society. This database captures up to 70% of nation's cancer cases through its participating hospitals.¹⁷ It offers vital status from 1998 to 2006, the years of diagnosis for patients included in this study, and provides a comprehensive list of all of the variables discussed so far that contribute to OS. Based on a review of relevant literature, the NCDB has not been used in OS calculators for colon cancer. The purpose of this study was to create a user-friendly postoperative OS calculator for patients with stage II/III colon cancer derived from the NCDB and develop it into a point-of-care mobile application. The benefits of this tool include providing an individualized estimate of OS after surgery, which has practical value for the patient, and quantifying the predicted effect of adjuvant treatment on postoperative OS to assist in the decision-making process for further treatment.

METHODS

Patients

We queried the American College of Surgeons National Cancer Data Base (ACS-NCDB), 2004 to 2006, to identify all patients with invasive colon adenocarcinoma, pathologic stages II and III, who had undergone surgery. Institutional review was not required for this study. Although data are available from 1998 to 2011, this shorter time period was chosen because the NCDB began recording site-specific factors in 2004 (which includes items like CEA) and also because vital status was validated only until 2006 at the time of this study. Patients

with adenocarcinoma were identified using International Statistical Classification of Diseases (ICD O-3) codes 8140 to 8148, 8200, 8260 to 8263, and 8480 to 8496.

Inclusion criteria consisted of patients with pathologic stage II or III colon adenocarcinoma who underwent surgical resection reported by the NCDB, recorded as local excision, partial colectomy, subtotal or hemicolectomy, total colectomy, total proctocolectomy, colectomy with resection of contiguous organs, and colon resection not otherwise specified. Patients with clinical or pathologic stage I or IV disease were excluded. Patients with malignant histology other than adenocarcinoma, carcinoma in situ, more than 1 recorded malignancy, or missing demographic or tumor characteristic data were also excluded from the analysis.

Patient factors reported in the NCDB include race, preoperative CEA levels, and Charlson-Deyo comorbidity score as a measure of comorbid conditions. Pathologic variables include site of tumor, grade, tumor size, pathologic TNM stage, number of nodes examined, number of positive nodes, and margin status. For years 2004 to 2006, the NCDB records TNM stage using the American Joint Commission on Cancer (AJCC) 6th edition.¹⁸ The length of postoperative hospital stay and unplanned readmission rates are also available through the NCDB. These items have been associated with postoperative morbidity and OS in patients with colon cancer.¹⁵

The NCDB also records data on the administration of adjuvant therapy (both chemotherapy and radiation), although the specific regimens are not available as part of the database. As reflected in the current guidelines from the National Comprehensive Cancer Network (NCCN), for stage II patients with high-risk features (such as advanced grade, depth of invasion or LVI/PNI) and for patients with stage III disease, adjuvant chemotherapy is often recommended.¹⁹ Adjuvant radiation is not typically a part of the treatment algorithm for colon cancer, but may be used for locally advanced tumors in stage II or III disease.

Statistical analysis

The pathologic stage II/III patients were randomly divided into testing (n_t) and validation (n_v) cohorts using an 80/20 allocation, respectively. Descriptive statistics were reported using the mean, median, and standard deviation for continuous variables, and frequencies and relative frequencies for categorical variables. Cohorts were compared using Mann-Whitney U and chi-square tests for continuous and categorical variables, respectively. Kaplan-Meier methods were used to summarize OS, from which estimates of median and 5-year survival were obtained with 95% CIs. Comparisons of OS were made using the log-rank test.

The following analyses were conducted using the n_t cohort. Univariate associations between OS and patient variables were examined using a Cox regression model from which hazard ratios (HRs) were obtained with 95% CI. A multivariable Cox regression model was obtained using a bootstrap backwards selection method. The backwards selection method was applied to 500 bootstrap resamples, and variables that were selected at least 70% of the time were included in the final model. Candidate variables for the multivariable model are represented in Table 1. The final model was fit using Firth's penalized function. Firth's function is a correction method for aiding in convergence of the log likelihood associated

with maximum likelihood methods.^{20,21} It is particularly useful in situations with a low number of events and a higher number of covariates. Hazard ratios with 95% CIs were obtained from parameter estimates. The baseline 5-year OS rate was obtained from the fitted Cox regression model using the BASELINE function available in SAS v 9.4 software's PHREG procedure, and corresponds to the estimated 5-year OS rate when all covariates are set to the reference category or zero, as appropriate. The parameter estimates and estimated baseline 5-year OS rate were used to generate prediction models for 5-year OS. The model was recalibrated using standard bootstrap cross-validation methods. Model performance was assessed using 5-year receiver operating characteristic (ROC) curves, the corresponding area under the ROC curve (AUC), and calibration plots.

The 5-year OS prediction models were then applied to the n_v cohort, with model performance assessed using the 5-year ROC curve with corresponding AUC, and calibration plots. The calibration plots for 5-year OS plot the predicted survival rates on the horizontal axis vs the observed survival rates on the vertical axis. The prediction models were applied to the data and predicted 5-year OS probabilities were obtained. Based on these probabilities, the patients were then divided into risk vigintiles (n_t cohort) or deciles (n_v cohort). The 5-year OS rates were then estimated for each risk group using standard Kaplan-Meier methods. These calibration plots identify when the prediction models over, under, or accurately predicted the 5-year OS rates.

All analyses were conducted using SAS v 9.4, at a significance (p value) level of 0.05.

Mobile application software design

With the assistance of Bitwise Analytics, Inc, the mathematical survival prediction equation calculating 5-year OS for postoperative patients with stage II/III colon cancer was incorporated into a web-based mobile application. This technologically savvy medium can function on smart-phones, tablets, and other portable electronic devices.

RESULTS

Data extraction from 2004 to 2006 yielded 230,520 patients. Patients with nonadenocarcinoma histology (11,039), carcinoma in situ (14,034), or more than 1 recorded malignancy or an unknown number of malignancies (53,156), were sequentially excluded, leaving 152,291 patients. Of these, a total of 129,040 patients underwent surgical resection; local endoscopic or ablative procedures were excluded. Patients with stage I or IV disease, as well as those with unrecorded stage, were also excluded (62,240). Last, after exclusion of patients with greater than 10% missing demographic or tumor-specific data, a total of 34,176 patients were included in the analysis. Among these patients, a total of 18,006 (52.7%) patients had pathologic stage II disease, and 16,170 (47.3%) patients had stage III disease. As described in the methods, 27,341 patients (80.0%) and 6,835 patients (20.0%) were then randomly divided into n_t and n_v cohorts, respectively.

Patient and tumor variables used to generate the OS calculator are shown in Table 1 with regard to each variable's amount of missing data. After allocation to the n_t and n_v cohorts, the comparison of n_t and n_v cohorts showed no statistically meaningful differences (Table 1).

The overall patient and tumor characteristics are shown in Table 1 under the column designated “Overall.”

Figure 1 shows the standard Kaplan-Meier curves for OS. Median follow-up was 75.0 months (range 0.0 to 111.1 months). The median OS was 91.6 months, with a 5-year OS of 0.63 (95% CI 0.62 to 0.63).

As a result of the large sample size, univariate prediction modeling using n_t showed that the majority of variables were significantly associated with 5-year OS (Table 2). Although tumor site was statistically significant, it was not included in the testing cohort for several reasons. First, the effect of different tumor sites on OS was not large compared to the other variables, with the largest effect being an HR equal to 1.2 (sigmoid compared with reference, ie, appendix). Therefore, in general, tumor site was deemed to have a small effect on the model. Second, because surgical site is reflected in the specific type of operation performed, there would be a redundancy in this postoperative prediction model and in turn, the mobile application. In other words, the type of operation intrinsically captures the tumor site. Multivariable analysis is shown in Table 3. Each of these variables was independently associated with OS to significant levels.

Performance of the model derived from the n_t cohort is reflected in the receiver operating characteristic curve (ROC) and the area under the ROC curve (AUC), as shown in Figure 2. The AUC equals 0.779, representing good predictive performance of this model. The mathematical model for predicted probability “P” of 5-year OS is expressed as follows: $P(5\text{-year OS}) = 0.987 \cdot [\exp(I_p)]$, where the equation for I_p is shown in Figure 3.

The recalibrated model constructed by using the n_v cohort showed a slight drop in performance, with $AUC = 0.772$ (Fig. 4), such that the prediction model consistently underestimates 5-year OS. An AUC of 0.772 still indicates strong performance,^{22–24} showing that this model has appropriate utility in identifying the factors associated with risk for OS, and therefore provides an estimate of 5-year OS.

Mobile application software

The above mathematical equation was used to create software suited for portable electronic devices, culminating in a working prototype of a novel web-based calculator application. Data can be entered in a user-friendly manner to estimate the probability of 5-year OS, which is instantly displayed on the screen. An example of calculator use is shown Figure 5. Panel A shows the different variables entered for a hypothetical patient. Panel B shows the generated 5-year OS for these individualized variables.

DISCUSSION

For patients who undergo surgery for colon cancer, postoperative survival calculators can provide useful information and aid in the care of these patients. In this study, a new mobile application was developed to estimate postoperative 5-year OS in patients treated for stage II/III colon cancer. The mathematical prediction equations integrated into this mobile application incorporate a comprehensive list of patient demographic variables, pathologic

characteristics, postoperative factors (such as unplanned readmission), and adjuvant treatment, which have all been shown to affect OS in colon cancer. Most of these factors are expressed as discrete or nominal variables within the equation. For example, female sex would receive a value of 1 (or “yes”), and male sex would receive a value of 0 (or “no”). Four of the variables are expressed in a continuous fashion and include age (in years), number of positive lymph nodes, number of lymph nodes examined, and inpatient stay (in days). The coefficient preceding each variable reflects the weight of that specific variable on the OS prediction. Larger coefficients, such as those associated with adjuvant chemotherapy, Charlson-Deyo score, pathologic stage, surgery type, or margin status, were found to have a greater impact on the prediction model.

The “Ip” component of the mathematical equation is an exponent for the base constant 0.987. Consequently, the larger the final value of Ip, the lower the OS and vice-versa. Therefore, variables that are additive to Ip subsequently lessen the OS risk. These variables include older age, male sex, Native American or black race, higher Charlson-Deyo score, type of surgery, worse differentiation, larger size, advanced stage, margin positivity, prolonged hospital stay, unplanned readmission, adjuvant radiation, and elevated CEA level. Conversely, variables that are subtractive to Ip result in elimination of the variable (through a value of 0) or decreased Ip, which subsequently increases the OS. These variables include younger age, female sex, white or Asian race, Charlson-Deyo score of 0, well-differentiated grade, smaller size, earlier stage, negative margins, shorter hospital stay, planned readmission, and adjuvant chemotherapy. In sum, this novel equation correctly accounts for the contribution of each of these variables in the model in the expected clinical manner and assigns the appropriate weight to each variable as derived from the NCDB, resulting in an accurate prediction calculator for OS (AUC = 0.772).

Based on the current literature, the NCDB colon cancer database has not been used to generate survival calculators. The items included in the database are comprehensive of many patient and oncologic factors, which allows for incorporation of a wide range of variables known to influence OS into the prediction model. Similar to existing colon cancer survival calculators,^{9,11} the impact on OS from tumor grade, tumor size, pathologic TNM stage, and node status are represented in this mobile application. Patient comorbidity profile has been shown to affect colon cancer survival.^{25,26} Some existing calculators use a measure of functional or comorbid status, while others do not. Different from existing calculators, the influence of postoperative short-term outcomes was also integrated into OS prediction. Prolonged hospital length of stay and unplanned readmission related to postoperative complications after surgical resection have been reported to have an impact on OS.¹⁵ When predicting OS, it is not only important to communicate to the patient the individualized risk of death as a result from cancer, but also to account for competing causes of death in order to provide a more meaningful framework of the OS prediction. This is a well-established concept in communicating OS to the patient,²⁷ and this novel mobile application accounts for a wide range of important information related to OS prediction by including comorbid status and postoperative clinical factors. By quantifying the effect of postoperative length of stay on OS, this calculator may also serve as a useful tool for surgical quality metrics.

One of the important benefits of the NCDB is that it provides a large amount of patient data (from nearly 35,000 patients), which allows for a robust statistical analysis. One calculator derived from a large institutional database included approximately 1,300 patients.¹⁰ Moreover, online prediction calculators derived from institutional databases are not as comprehensive in the factors included in their models. This may limit the applicability to colon cancer patients in general across the nation. Other calculators, such as the Adjuvant! Online calculator, have been derived using large nationwide patient samples from the SEER database.^{9,11} The SEER database offers a robust sample size and many of the oncologic variables associated with cancer survival. It also provides information about patient comorbidities and has the additional benefit of providing some information on the type of adjuvant chemotherapy given. However, it does not incorporate postoperative complications. Similarly, the on-line calculator derived from the ACCENT database provides some additional granularity on adjuvant chemotherapy, but again, does not factor in postoperative complications. Therefore, this calculator serves as a progressive step in the line of other risk calculators available for surgical patients. One of the future directions includes using this innovative mobile application prospectively and assessing its validity in our colorectal clinic. Another area of interest will be to compare our calculator with the other existing calculators discussed earlier in terms of prediction performance.

Another benefit of this mobile application is that it provides additional, individualized point-of-care estimation of the benefit of receiving adjuvant treatments, both chemotherapy and radiation, for patients with stage II/III disease. The impact on OS for adjuvant chemotherapy can be readily incorporated into the OS prediction and compared with the OS predicted without adjuvant chemotherapy. This information can contribute to other existing calculators that estimate the benefit of adjuvant chemotherapy.^{12,28} This may assist in the decision making process of patients and their caregivers as to whether to receive adjuvant chemotherapy after surgery, particularly in the context of other patient-specific factors such as age, comorbidities, and the presence of postoperative complications (as reflected by prolonged hospital stay and unplanned readmission). In contrast, receiving adjuvant radiation therapy is a negative contributor to OS. This was expected because adjuvant radiation is not a typical component of multimodal therapy for colon cancer, but is more often used in patients with locally aggressive cancer, which is a poor prognostic factor. The need for adjuvant chemotherapy or radiation should be assessed in conjunction with clinical judgment. For example, predictions of OS with adjuvant therapy will be less relevant for patients with stage II disease with low risk features compared with patients with stage III disease.

Limitations of this calculator include those that are inherent to large databases, which are often less granular, specifically in terms of the particular types of adjuvant therapy received or details regarding the circumstances contributing to extended lengths of postoperative hospital stay or unplanned readmissions. The presence of LVI or PNI on pathologic specimens has been shown to affect prognosis.^{4,8} Long-term disease-specific and recurrence-free survivals in patients with either LVI or PNI have been shown to be half those of patients without these characteristics.^{4,29} Although it was preferred to incorporate these oncologic factors into the prediction models, the NCDB began recording LVI status in 2010. Therefore, LVI status could not be evaluated with respect to its association with OS.

Regarding PNI, the NCDB began recording this in 2004, but there was an overwhelming amount of missing data. Therefore, PNI was not incorporated into this calculator. Nationwide databases, including the NCDB, are always in a phase of evolution and constantly maturing. Additional data will likely be released and available in the future, which will allow an update and recalibration of the prediction modeling processes.

In this model, patients with more than 10% missing data were excluded from the testing cohort. Although there are methods to incorporate patients with missing data into the model through a variety of different imputational techniques, these methods are not without their own limitations. Exclusion of these patients still allowed for a robust sample size of more than 34,000 patients. Within this sample, there was a high percentage of complete tumor-specific information as well as demographic information, with the most missing data related to ethnicity. However, for the variables associated with high coefficients in the prediction model equation, there was near-complete or complete data available, which is again reflected in the strong performance of the model (AUC = 0.772).

CONCLUSIONS

In conclusion, this study reports the development of an individualized, postoperative OS calculator for colon cancer integrated into a mobile-based application platform. Like other currently available calculators, this novel application is not intended to obviate clinical judgment when estimating individual OS or advising for adjuvant treatment. For example, as previously discussed, the extent to which patients with high risk stage II colon cancer are represented in the model is unknown, so predictions of OS may be overestimated in this group of patients. The use of adjuvant radiation for patients with locally advanced tumors may offer improved local control. However, as shown in this model, the use of adjuvant radiation decreases OS estimates. Although adjuvant radiation portends worse OS as the result of advanced disease, it should still be considered as part of the treatment regimen. Therefore, this application is meant to be used as a supplemental tool by the health care provider during the patient encounter. Despite these limitations, this prediction model is unique to existing calculators in that by using large nationwide data, it incorporates a diverse set of patient-specific, tumor-related, and outcome-dependent variables, culminating in a highly comprehensive and statistically robust prediction calculator.

Acknowledgments

We thank Bitwise Analytics, Inc. for translating our colon cancer survival calculator into a web-based mobile application. We also thank the American College of Surgeons Commission on Cancer for providing access to the Participant User File from the National Cancer Data Base. We intend to make the mobile application freely available online.

Abbreviations and Acronyms

AUC	area under the curve
HR	hazard ratio
LVI	lymphovascular invasion

NCDB	National Cancer Data Base
n_t	testing
n_v	validation
OS	overall survival
PNI	perineural invasion
ROC	receiver operating characteristic
SEER	Surveillance Epidemiology, and End Results

References

1. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J National Cancer Inst.* 2004; 96:1420–1425.
2. Edge, S., Byrd, DR., Compton, et al. *AJCC Cancer Staging Manual*. 7th. New York: Springer-Verlag; 2010.
3. Saha S, Shaik M, Johnston G, et al. Tumor size predicts long-term survival in colon cancer: an analysis of the National Cancer Data Base. *Am J Surg.* 2014; 209:570–574. [PubMed: 25601557]
4. Liebig C, Ayala G, Wilks J, et al. Perineural invasion is an independent predictor of outcome in colorectal cancer. *J Clin Oncol.* 2009; 27:5131–5137. [PubMed: 19738119]
5. Chandrasinghe PC, Ediriweera DS, Hewavisenthi J, et al. The total number of lymph nodes harvested is associated with better survival in stages II and III colorectal cancer. *Indian J Gastroenterol.* 2014; 33:249–253. [PubMed: 24048680]
6. Andre T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *N Engl J Med.* 2004; 350:2343–2351. [PubMed: 15175436]
7. Andre T, Boni C, Navarro M, et al. Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. *J Clin Oncol.* 2009; 27:3109–3116. [PubMed: 19451431]
8. Fujita S, Shimoda T, Yoshimura K, et al. Prospective evaluation of prognostic factors in patients with colorectal cancer undergoing curative resection. *J Surg Oncol.* 2003; 84:127–131. [PubMed: 14598355]
9. Weiser MR, Gonen M, Chou JF, et al. Predicting survival after curative colectomy for cancer: individualizing colon cancer staging. *J Clin Oncol.* 2011; 29:4796–4802. [PubMed: 22084366]
10. Weiser MR, Landmann RG, Kattan MW, et al. Individualized prediction of colon cancer recurrence using a nomogram. *J Clin Oncol.* 2008; 26:380–385. [PubMed: 18202413]
11. Chang GJ, Hu CY, Eng C, et al. Practical application of a calculator for conditional survival in colon cancer. *J Clin Oncol.* 2009; 27:5938–5943. [PubMed: 19805670]
12. Renfro LA, Grothey A, Xue Y, et al. ACCENT-based web calculators to predict recurrence and overall survival in stage III colon cancer. *J National Cancer Inst.* 2014; 106:1–9.
13. O'Connell MJ, Campbell ME, Goldberg RM, et al. Survival following recurrence in stage II and III colon cancer: findings from the ACCENT data set. *J Clin Oncol.* 2008; 26:2336–2341. [PubMed: 18467725]
14. Shibutani M, Maeda K, Nagahara H, et al. Significance of CEA and CA19-9 combination as a prognostic indicator and for recurrence monitoring in patients with stage II colorectal cancer. *Anticancer Res.* 2014; 34:3753–3758. [PubMed: 24982398]
15. Schneider EB, Hyder O, Brooke BS, et al. Patient readmission and mortality after colorectal surgery for colon cancer: impact of length of stay relative to other clinical factors. *J Am Coll Surg.* 2012; 214:390–398. discussion 398–399. [PubMed: 22289517]

16. Park IJ, Choi GS, Lim KH, et al. Serum carcinoembryonic antigen monitoring after curative resection for colorectal cancer: clinical significance of the preoperative level. *Ann Surg Oncol*. 2009; 16:3087–3093. [PubMed: 19629600]
17. Winchester DP, Stewart AK, Phillips JL, et al. The National Cancer Data Base: past, present, and future. *Ann Surg Oncol*. 2010; 17:4–7. [PubMed: 19847564]
18. American Joint Committee on Cancer. Manual for Staging of Cancer. 6th. Philadelphia: Lippincott Raven Publishers; 2002.
19. National Comprehensive Cancer Network. Colon Cancer (Version 4.2014). Available at: http://www.nccn.org/professionals/physician_gls/pdf/colon.pdf. Accessed August 1, 2015
20. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993; 80:27–38.
21. Heinze G, Schemper M. A solution to the problem of monotone likelihood in Cox regression. *Biometrics*. 2001; 57:114–119. [PubMed: 11252585]
22. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. *CJEM*. 2006; 8:19–20. [PubMed: 17175625]
23. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005; 61:92–105. [PubMed: 15737082]
24. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]
25. McKay A, Donaleshen J, Helewa RM, et al. Does young age influence the prognosis of colorectal cancer: a population-based analysis. *World J Surg Oncol*. 2014; 12:370. [PubMed: 25466394]
26. Sogaard M, Thomsen RW, Bossen KS, et al. The impact of comorbidity on cancer survival: a review. *Clin Epidemiol*. 2013; 5:3–29. [PubMed: 24227920]
27. Howlader N, Mariotto AB, Woloshin S, et al. Providing clinicians and patients with actual prognosis: cancer in the context of competing causes of death. *J National Cancer Inst Monographs*. 2014; 2014:255–264. [PubMed: 25417239]
28. Bardia A, Loprinzi C, Grothey A, et al. Adjuvant chemotherapy for resected stage II and III colon cancer: comparison of two widely used prognostic calculators. *Semin Oncol*. 2010; 37:39–46. [PubMed: 20172363]
29. Nissan A, Stojadinovic A, Shia J, et al. Predictors of recurrence in patients with T2 and early T3, N0 adenocarcinoma of the rectum treated by surgery alone. *J Clin Oncol*. 2006; 24:4078–4084. [PubMed: 16943525]

Cohort	# At risk Baseline	# At Risk 5 years	# Deaths 5 years	# Censored (lost to follow-up)
Testing	27,341	13,664	9,491	4,186
Validation	6,835	3,470	2,343	1,022

Figure 1.

Overall survival (OS) used to create calculator. No difference between the testing and validation cohorts was observed, with similar median OS of the testing cohort (91.5 months) compared with the validation cohort (91.9 months).

Figure 2.

(A) Calibration plot and (B) receiver operating characteristic curve for the building prediction model using the testing cohort. In the calibration plot, proximity of the dotted line to the solid line indicates performance of the model, whereby the closer the 2 lines are, the higher the accuracy of the predicted 5-year overall survival to the observed or actual 5-year overall survival. Bars indicate 95% CIs about a given data point. In the receiver operating characteristic curve, the area under the curve (AUC) equals 0.779. The closer this value is to 1.0, the higher the performance of the model.

$$\begin{aligned}
 Ip = & 0.99981 \times [0.03816 \times \text{age in years} - 0.20852 \times (\text{Sex} = \text{female}) - 0.16849 \times (\text{Race} = \text{other}) - \\
 & 0.28037 \times (\text{Race} = \text{Asian}) + 0.26682 \times (\text{Race} = \text{Native American}) + 0.19286 \times (\text{Race} = \\
 & \text{Black}) + 0.55091 \times (\text{CD} = 2) + 0.23901 \times (\text{CD} = 1) + 0.41081 \times (\text{Size} > 9 \text{ cm}) + 0.30950 \times \\
 & (4 < \text{Size} \leq 9 \text{ cm}) + 0.25587 \times (1 < \text{Size} \leq 4 \text{ cm}) + 0.36650 \times (\text{colectomy NOS}) + 0.54036 \times \\
 & (\text{colectomy or coloproctectomy}) + 0.57692 \times (\text{total proctocolectomy}) + 0.51723 \times (\text{total} \\
 & \text{colectomy}) + 0.36367 \times (\text{subtotal or hemicolectomy}) + 0.31328 \times (\text{partial colectomy}) + \\
 & 0.01919 \times \text{LOS} + 0.29753 \times (\text{Readmission} = \text{unplanned}) - 0.00667 \times (\text{Readmission} = \\
 & \text{planned}) + 0.31852 \times (\text{CEA} = \text{elevated}) - 0.69301 \times (\text{adjuvant chemotherapy} = \text{yes}) + \\
 & 0.38279 \times (\text{adjuvant radiation} = \text{yes}) + 0.07113 \times (\text{grade} = \text{poorly differentiated}) + 0.29814 \\
 & \times (\text{grade} = \text{undifferentiated}) + 0.53160 \times (\text{Margins} = \text{positive}) - 0.01886 \times (\text{number nodes} \\
 & \text{examined}) + 0.08380 \times (\text{number nodes positive}) + 0.49783 \times (\text{pathologic stage} = 3)]
 \end{aligned}$$

Figure 3.

Ip component of the 5-year overall survival prediction model. Parentheses indicate the variable incorporated into the model. For continuous variables (including age, length of stay, number of nodes examined, and number of nodes positive), the numeric value specific to the patient is entered. All other variables are treated as indicator variables, whereby the parentheses take on either a value of 1 if the patient has that characteristic or a value of zero if the patient does not. CD, Charlson-Deyo comorbidity score; CEA, carcinoembryonic antigen level; LOS, postoperative length of stay; NOS, not otherwise specified.

Figure 4.

(A) Calibration plot and (B) receiver operating characteristic curve for the prediction model after validation with the testing cohort. The area under the curve (AUC) shows similar performance to the building prediction model, with a value of 0.772.

Stage I/III Colon Cancer Survival Calculator

Age : 70

Gender : Female

Race : White

Charlson's Deyo Score : 0

Tumor Size: 4< and <=9 cm

Procedure: Hemi Colectomy or Subtotal Colect...

Length of Stay (# of Days) 4

Readmission: None

CEA: Normal

Tumor Grade: Moderately Differentiated

Pathologic Stage: Stage 2

Margins: Negative

Lymph Nodes Examined (#) 25

Lymph Nodes Positive (#)

Adjuvant Chemotherapy: Yes

Adjuvant Radiation: No

Clear Calculate

Stage I/III Colon Cancer Survival C...

Length of Stay (# of Days) 4

Readmission: None

CEA: Normal

Tumor Grade: Well Differe...

Result

5 Year Survival(%) : 82.92

OK

Lymph Nodes Positive (#) 4

Adjuvant Chemotherapy: Yes

Adjuvant Radiation: Yes

Clear Calculate

A
B

Figure 5. Mobile application prototype. (A) Example of individual demographic, tumor-specific, and postoperative clinic characteristics entered into the calculator. (B) The generated estimate of 5-year overall survival.

Table 1

Characteristics of Patients Used to Construct the Colon Cancer Postoperative Overall Survival Calculator

Characteristic	Testing (n _t) cohort	Validation (n _v) cohort	Overall	p Value
Overall, n (%)	27,341 (80.0)	6,835 (20.0)	34,176 (100)	
Age, y, mean (SD)	68.8 (13.7)	68.5 (13.8)	68.8 (13.7)	0.15
Sex, n (%)				
Male	12,471 (45.6)	3,118 (45.6)	15,589 (45.6)	0.99
Female	14,870 (54.4)	3,717 (54.4)	18,587 (54.4)	
Race, n (%) [*]				
White	23,299 (85.9)	5,750 (84.8)	29,049 (85.7)	0.16
Black	2,953 (10.9)	804 (11.9)	3,757 (11.1)	
Native American	52 (0.2)	15 (0.2)	67 (0.2)	
Asian	694 (2.6)	177 (2.6)	871 (2.6)	
Other	127 (0.5)	38 (0.6)	165 (0.5)	
Hispanic origin, n (%) [*]				
Non-Hispanic	23,358 (94.9)	5,860 (94.9)	29,218 (94.9)	0.96
Hispanic origin	1,251 (5.1)	315 (5.1)	1,566 (5.1)	
Charlson-Deyo, n (%) [†]				
0	18,761 (68.6)	4,707 (68.9)	23,468 (68.7)	0.89
1	6,321 (23.1)	1,561 (22.8)	7,882 (23.1)	
2	2,259 (8.3)	567 (8.3)	2,826 (8.3)	
Primary site, n (%)				
Appendix	174 (0.6)	58 (0.8)	232 (0.7)	0.15
Cecum	6,484 (23.7)	1,661 (24.3)	8,145 (23.8)	
Ascending colon	5,882 (21.5)	1,486 (21.7)	7,368 (21.6)	
Hepatic flexure	4,182 (15.3)	1,098 (16.1)	5,280 (15.4)	
Transverse colon	1,064 (3.9)	256 (3.7)	1,320 (3.9)	
Splenic flexure	1,637 (6.0)	402 (5.9)	2,039 (6.0)	
Descending colon	7,203 (26.3)	1,715 (25.1)	8,918 (26.1)	
Sigmoid	345 (1.3)	76 (1.1)	421 (1.2)	
Overlapping	370 (1.4)	83 (1.2)	453 (1.3)	
Grade, n (%) [*]				
Well	2,160 (8.1)	563 (8.4)	2,723 (8.1)	0.65
Moderate	18,614 (69.4)	4,664 (69.5)	23,278 (69.5)	
Poor	5,654 (21.1)	1,399 (20.8)	7,053 (21.0)	
Undifferentiated	377 (1.4)	85 (1.3)	462 (1.4)	
Tumor size, cm, n (%) [*]				
<1	171 (0.7)	42 (0.6)	213 (0.7)	0.90
1–4	10,579 (40.5)	2,671 (41.0)	13,250 (40.6)	

Characteristic	Testing (n _t) cohort	Validation (n _v) cohort	Overall	p Value
4-9	13,781 (52.8)	3,405 (52.3)	17,186 (52.7)	
>9	1,580 (6.1)	392 (6.0)	1,972 (6.0)	
Positive nodes, n, mean (SD) *	1.8 (3.3)	1.8 (3.5)	1.8/3.3	0.69
Nodes examined, n, mean (SD) *	16.5 (10.0)	16.5 (10.0)	16.5/10.0	0.96
Type of operation, n (%)				
Local tumor excision	71 (0.3)	10 (0.1)	81 (0.2)	0.41
Partial colectomy	8,288 (30.3)	2,076 (30.4)	10,364 (30.3)	
Subtotal or hemicolectomy	16,127 (59.0)	3,994 (58.4)	20,121 (58.9)	
Total colectomy	397 (1.5)	99 (1.4)	496 (1.5)	
Total proctocolectomy	51 (0.2)	12 (0.2)	63 (0.2)	
Colectomy or coloproctectomy	2,197 (8.0)	582 (8.5)	2,779 (8.1)	
Colectomy NOS	210 (0.8)	62 (0.9)	272 (0.8)	
Margins, n (%) *				
Negative	25,548 (94.4)	6,392 (94.4)	31,940 (94.4)	0.89
Positive	1,506 (5.6)	380 (5.6)	1,886 (5.6)	
Inpatient stay, d, mean (SD) *	7.6 (6.8)	7.5 (6.1)	7.6 (6.7)	0.98
Readmission, n (%) *				
None	23,794 (90.3)	5,972 (90.3)	29,766 (90.3)	0.79
Planned	1,085 (4.1)	261 (3.9)	1,346 (4.1)	
Unplanned	1,478 (5.6)	377 (5.7)	1,855 (5.6)	
Radiation, n (%) *				
No	26,500 (97.9)	6,654 (98.3)	33,154 (98.0)	0.052
Yes	555 (2.1)	114 (1.7)	669 (2.0)	
Chemotherapy, n (%) *				
No	15,520 (59.5)	3,872 (59.2)	19,392 (59.5)	0.69
Yes	10,557 (40.5)	2,664 (40.8)	13,221 (40.5)	
CEA, n (%)				
Negative	16,207 (59.3)	4,092 (59.9)	20,299 (59.4)	0.37
Positive	11,134 (40.7)	2,743 (40.1)	13,877 (40.6)	
TNM pathologic stage, n (%)				
II	14,402 (52.7)	3,604 (52.7)	18,006 (52.7)	0.94
III	12,939 (47.3)	3,231 (47.3)	16,170 (47.3)	

Patients diagnosed between 2004 and 2006 were identified from the NCDB. From the entire cohort, 80% were randomly assigned to model development derived from the n_t cohort (designated building [testing] cohort) and 20% to the n_v cohort.

* Variables with less than 100% available data from the NCDB, years 2004 to 2006. Missing data: race 0.8%, Hispanic origin 9.9%, grade 1.9%, tumor size 4.5%, number positive nodes 0.5%, number nodes examined 0.6%, margins 1.0%, inpatient stay 6.5%, readmission 3.5%, radiation 1.1%, chemotherapy 4.6%.

[†]Charlson-Deyo comorbidity score is an estimate of comorbid conditions based on ICD-9 diagnosis codes. A score of 0 indicates no comorbidities. Point values are assigned to comorbid conditions based on severity. The NCDB truncates possible scores to 0, 1, and 2 due to the small proportion of cases exceeding a score of 2.

NCDB, National Cancer Data Base; NOS, not otherwise specified.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Univariate Analysis of Variables Associated with Overall Survival, Showing the Kaplan-Meier Survival Estimates of the Survival Rates and the Cox Proportional Hazards Model Hazards Ratios, 95% CI, and Associated p Values

Variable	Median survival, mo (95% CI)	5-y Survival rate, % (95% CI)	Hazard ratio (95% CI)	p Value
Age, y			1.047 (1.046, 1.049)	< 0.001
Sex				
Male (reference)	90.5 (87.4, 93.4)	62.2 (61.3, 63.1)	1.00	0.40
Female	92.4 (89.9, 95.1)	62.9 (62.1, 63.7)	0.98 (0.95, 1.02)	
Race				
White (reference)	89.9 (88.3, 92.1)	62.2 (61.6, 62.9)	1.00	< 0.001
Black	97.1 (89.9, 100.7)	62.1 (60.2, 63.9)	0.98 (0.92, 1.04)	
Native American	79.2 (63.7, NR *)	69.6 (53.9, 80.9)	0.89 (0.57, 1.40)	
Asian	NR (NR, NR) *	71.8 (68.0, 75.1)	0.67 (0.58, 0.77)	
Other	NR (77.9, NR *)	70.0 (60.4, 77.7)	0.78 (0.57, 1.06)	
Hispanic origin				
Non-Hispanic (reference)	90.3 (88.5, 92.5)	62.2 (61.5, 62.8)	1.00	< 0.001
Hispanic origin	NR (97.7, NR *)	67.5 (64.6, 70.2)	0.81 (0.74, 0.90)	
Charlson-Deyo				
0 (reference)	103.4 (101.1, NR *)	68.0 (67.3, 68.7)	1.00	< 0.001
1	71.9 (69.5, 74.9)	55.0 (53.8, 56.3)	1.56 (1.49, 1.62)	
2	42.6 (38.0, 46.1)	39.5 (37.4, 41.6)	2.45 (2.32, 2.60)	
Primary site				
Appendix (reference)	100.8 (65.3, NR *)	61.2 (53.1, 68.3)	1.00	< 0.001
Cecum	82.3 (79.3, 85.3)	58.9 (57.7, 60.2)	1.16 (0.92, 1.47)	
Ascending colon	87.2 (83.5, 90.8)	62.0 (60.7, 63.3)	1.08 (0.85, 1.36)	
Hepatic flexure	88.7 (84.5, 92.7)	60.5 (59.0, 62.1)	1.12 (0.87, 1.41)	
Transverse colon	NR * (92.3, NR *)	66.4% (63.3, 69.3)	0.89 (0.69, 1.14)	
Splenic flexure	105.0 (100.7, NR *)	65.6 (63.1, 67.9)	0.91 (0.71, 1.16)	
Descending colon	101.5 (98.6, 106.0)	66.9 (65.8, 68.0)	0.88 (0.70, 1.12)	
Sigmoid	88.5 (65.5, NR *)	56.3 (50.7, 61.6)	1.20 (0.91, 1.59)	
Overlapping	79.4 (64.4, NR *)	57.7 (52.1, 62.9)	1.16 (0.88, 1.53)	
Grade				
Well differentiated (reference)	100.6 (94.2, NR *)	69.4 (67.3, 71.3)	1.00	< 0.001
Moderate	95.3 (93.3, 97.4)	65.1 (64.4, 65.9)	1.11 (1.03, 1.20)	
Poor/undifferentiated	67.0 (63.2, 71.1)	52.2 (50.9, 53.5)	1.65 (1.52, 1.78)	
Tumor size, cm				

Variable	Median survival, mo (95% CI)	5-y Survival rate, % (95% CI)	Hazard ratio (95% CI)	p Value
<1 (reference)	NR (96.3, NR [*])	75.3 (67.8, 81.2)	1.00	< 0.001
1–4	93.4 (90.9, 96.8)	64.6 (63.6, 65.5)	1.50 (1.14, 1.98)	
4–9	89.6 (86.8, 92.3)	61.3 (60.5, 62.2)	1.65 (1.25, 2.17)	
>9	83.6 (73.2, 92.9)	57.7 (55.1, 60.2)	1.87 (1.41, 2.49)	
No. positive nodes, 1 node increase			1.075 (1.070, 1.079)	< 0.001
No. nodes examined, 1 node increase			0.982 (0.980, 0.984)	< 0.001
Type of operation				
Local tumor excision (reference)	NR (86.1, NR)	69.2 (56.4, 79.0)	1.00	< 0.001
Partial colectomy	99.9 (95.1, 103.4)	65.6 (64.5, 66.6)	1.13 (0.75, 1.68)	
Subtotal or hemicolectomy	89.6 (87.2, 92.3)	61.9 (61.1, 62.7)	1.29 (0.86, 1.93)	
Total colectomy	99.4 (72.7, NR [*])	58.7 (53.5, 63.6)	1.36 (0.89, 2.09)	
Total proctocolectomy	97.7 (NR, NR [*])	66.6 (51.3, 78.1)	0.97 (0.52, 1.80)	
Colectomy or coloproctectomy	78.5 (70.9, 84.5)	56.5 (54.3, 58.6)	1.51 (1.01, 2.26)	
Colectomy NOS	93.6 (85.6, NR)	65.7 (58.4, 71.9)	1.08 (0.68, 1.71)	
Margins				
Negative (reference)	94.2 (92.3, 96.0)	64.1 (63.5, 64.7)	1.00	< 0.001
Positive	35.4 (31.8, 40.9)	39.4 (36.8, 42.0)	2.08 (1.94, 2.22)	
Inpatient stay, 1 d increase			1.025 (1.024, 1.026)	< 0.001
Readmission				
None (reference)	92.9 (90.9, 95.2)	63.0 (62.4, 63.7)	1.00	< 0.001
Planned	100.6 (94.2, NR [*])	68.6 (65.6, 71.3)	0.82 (0.74, 0.90)	
Unplanned	61.7 (53.7, 68.7)	50.5 (47.8, 53.1)	1.52 (1.42, 1.64)	
Radiation				
No (reference)	90.5 (88.9, 92.6)	62.4 (61.8, 63.0)	1.00	0.11
Yes	80.8 (65.6, 96.0)	56.5 (52.1, 60.7)	1.11 (0.98, 1.25)	
Chemotherapy				
No (reference)	71.3 (69.4, 73.0)	55.2 (54.3, 56.0)	1.00	< 0.001
Yes	NR (NR, NR) [*]	72.1 (71.2, 72.9)	0.52 (0.50, 0.54)	
CEA status				
Negative	101.5 (99.4, 103.6)	68.4 (67.7, 69.2)	1.00	< 0.001
Positive	70.5 (67.8, 73.0)	54.1 (53.1, 55.1)	1.57 (1.51, 1.63)	
TNM pathologic stage				
II (reference)	100.5 (97.9, 103.4)	68.6 (67.8, 69.4)	1.00	< 0.001
III	76.7 (73.9, 79.5)	56.0 (55.1, 56.9)	1.46 (1.41, 1.51)	

^{*} NR indicates that the value was not reached. This is a function of the number of events for a given variable. NOS, not otherwise specified.

Table 3

Multivariable Analysis of Variables to Identify Factors Independently Associated with Overall Survival

Variable	Hazard ratio (95% CI)	p Value
Age, 1-year increase	1.039 (1.037, 1.041)	< 0.001
Sex, female vs male (reference)	0.81 (0.78, 0.85)	< 0.001
Race		
Other vs white (reference)	0.85 (0.58, 1.23)	< 0.001
Asian vs white	0.76 (0.65, 0.88)	
Native American vs white	1.31 (0.81, 2.12)	
Black vs white	1.21 (1.13, 1.30)	
Charlson-Deyo		
2 vs 0 (reference)	1.74 (1.63, 1.85)	< 0.001
1 vs 0	1.27 (1.21, 1.33)	
Grade		
Poorly/undifferentiated vs well differentiated (reference)	1.35 (1.23, 1.47)	< 0.001
Moderate vs well	1.07 (0.99, 1.17)	
Tumor size, cm		
> 9 vs < 1 (reference)	1.51 (1.09, 2.09)	< 0.001
4–9 vs <1	1.36 (1.00, 1.87)	
1–4 vs <1	1.29 (0.94, 1.77)	
No. positive nodes, 1 node increase	1.087 (1.081, 1.094)	< 0.001
No. nodes examined, 1 node increase	0.981 (0.979, 0.984)	< 0.001
Type of operation		
Colectomy NOS vs local tumor excision (reference)	1.44 (0.75, 2.79)	< 0.001
Colectomy or colectectomy vs local tumor excision	1.72 (0.93, 3.16)	
Total proctocolectomy vs local tumor excision	1.78 (0.77, 4.12)	
Total colectomy vs local tumor excision	1.68 (0.89, 3.15)	
Subtotal or hemicolectomy vs local tumor excision	1.44 (0.78, 2.64)	
Partial colectomy vs local tumor excision	1.37 (0.75, 2.51)	
Margins, positive vs negative (reference)	1.70 (1.58, 1.84)	< 0.001
Inpatient stay, 1 d increase	1.019 (1.018, 1.021)	< 0.001
Readmission		
Unplanned vs none (reference)	1.35 (1.25, 1.45)	< 0.001
Planned vs none	0.99 (0.89, 1.11)	
Radiation, yes vs no (reference)	1.47 (1.27, 1.69)	< 0.001
Chemotherapy, yes vs no (reference)	0.50 (0.47, 0.53)	< 0.001
CEA, positive vs negative (reference)	1.38 (1.32, 1.43)	< 0.001
TNM pathologic stage, III vs II (reference)	1.65 (1.56, 1.73)	< 0.001

Data include patient demographics, tumor characteristics, postoperative factors, and adjuvant treatments for the 34,176 patients diagnosed with stage II or III colon cancer between 2004 and 2006 and treated with surgery. Each of these variables was incorporated into the prediction model in order to estimate 5-year overall survival.

NOS, not otherwise specified.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript