Ann Surg Oncol (2013) 20:161–174
DOI 10.1245/s10434-012-2555-4

Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ORIGINAL ARTICLE – COLORECTAL CANCER

Clinical Decision Support and Individualized Prediction of Survival in Colon Cancer: Bayesian Belief Network Model

Alexander Stojadinovic, MD[1,2,3], Anton Bilchik, MD, PhD[2,8,9], David Smith, PhD[4], John S. Eberhardt, BA[5], Elizabeth Ben Ward, MS[5], Aviram Nissan, MD[2,6], Eric K. Johnson, MD[1,2,10], Mladjan Protic, MD[2,11], George E. Peoples, MD[1,2,12], Itzhak Avital, MD[1,7], and Scott R. Steele, MD[1,2,10]

[1]Department of Surgery, Uniformed Services University of the Health Sciences, Bethesda, MD; [2]Clinical Trials Group, United States Military Cancer Institute, Washington, DC; [3]Division of Surgical Oncology, Department of Surgery, Walter Reed National Military Medical Center, Bethesda, MD; [4]Division of Biostatistics, City of Hope, Duarte, CA; [5]DecisionQ Corporation, Washington, DC; [6]Department of Surgery, Rabin Medical Center, Petah Tikvah, Israel; [7]Bon Secours Cancer Institute, Richmond, VA; [8]Department of Medicine, University of California Los Angeles, Los Angeles, CA; [9]California Oncology Research Institute, Santa Monica, California; [10]Department of Surgery, Madigan Army Medical Center, Tacoma, WA; [11]University of Novi Sad—Medical Faculty, Clinic of Abdominal, Endocrine and Transplantation Surgery, Clinical Centre of Vojvodina, Novi Sad, Serbia; [12]Department of Surgery, Brooke Army Medical Center, San Antonio, TX

ABSTRACT

Background. We used a large population-based data set to create a clinical decision support system (CDSS) for real-time estimation of overall survival (OS) among colon cancer (CC) patients. Patients with CC diagnosed between 1969 and 2006 were identified from the Surveillance Epidemiology and End Results (SEER) registry. Low- and high-risk cohorts were defined. The tenfold cross-validation assessed predictive utility of the machine-learned Bayesian belief network (ml-BBN) model for clinical decision support (CDS).

Methods. A data set consisting of 146,248 records was analyzed using ml-BBN models to provide CDS in estimating OS based on prognostic factors at 12-, 24-, 36-, and 60-month post-treatment follow-up.

Results. Independent prognostic factors in the ml-BBN model included age, race; primary tumor histology, grade and location; Number of primaries, AJCC T stage, N stage, and M stage. The ml-BBN model accurately estimated OS with area under the receiver-operating-characteristic curve of 0.85, thereby improving significantly upon existing AJCC stage-specific OS estimates. Significant differences in OS were found between low- and high-risk cohorts (odds ratios for mortality: 17.1, 16.3, 13.9, and 8.8 for 12-, 24-, 36-, and 60-month cohorts, respectively).

Conclusions. A CDSS was developed to provide individualized estimates of survival in CC. This ml-BBN model provides insights as to how disease-specific factors influence outcome. Time-dependent, individualized mortality risk assessments may inform treatment decisions and facilitate clinical trial design.

Colon cancer is the most common gastrointestinal tract malignancy in the United States, and it is the second leading cause of cancer-related mortality in the Western World.[1] This disease affects over a million people each year worldwide.[1] The primary treatment of non-metastatic colon cancer is surgical resection; however, disease in up to one-third of patients with node-negative disease undergoing potentially curative operation recurs.[1,2] Adjuvant systemic therapy is considered in high-risk node-negative and is indicated in node-positive patients with the specific aim of reducing disease recurrence and cause-specific mortality. As ethnic and socioeconomic disparities impact

Disclaimer The views expressed in this presentation are those of the authors and do not reflect the official policy of the Department of the Army, the Department of Defense, or the United States Government.

First Received: 1 May 2012;
Published Online: 17 August 2012

A. Stojadinovic, MD
e-mail: alexander.stojadinovic@med.navy.mil

oncological outcome in colon cancer, particularly among blacks and Hispanics, treatment-planning considerations expand beyond tumor stage-specific criteria.[3,4]

The current "gold standard" for defining disease extent and guiding treatment is the American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging system. This system is based on anatomic pathology, presently regarded as the most important determinant of oncological outcome.[5] Although the principal prognostic variables of this staging system are depth of primary tumor penetration (T stage) and number of regional nodes involved by tumor (N stage), the Hindgut Task Force in its revision of the sixth edition of the AJCC TNM staging system recognized the importance of: (1) the differences in prognosis amongst T4 subsets (visceral peritoneal involvement versus other organ adherence/invasion), (2) the prognostic impact of both number of tumor-involved nodes and total nodes assessed pathologically, (3) interplay between number of nodes involved and depth of tumor penetration of the colon wall or adjacent structures, and (4) the prognostic importance of satellite tumor deposits.[6] Although the evidence published to date does not support the use of molecular markers in the AJCC TNM staging system, the current edition also recognizes the heterogeneity of tumor biology and outcome within each stratum. This subcategorization within each stage stratum and categorical binning of continuous variables contribute to the limited predictive value of the AJCC TNM system in estimating overall survival, and the wide variability that may be demonstrated amongst patients within each AJCC stage.

An individual cancer patient's prognosis is typically estimated based on published consolidated data from large, heterogeneous study populations. Based on review of the published oncology literature, the Hindgut Task Force recognized important evidence—informed differences in oncological outcome amongst T4 tumors based on extent of disease. Further the Task Force noted: (1) worse prognosis among T1 and T2 tumors without regional nodal metastasis, but with satellite tumor deposits, (2) a significant influence of more penetrating tumors in stage II (node-negative) patients, and (3) a more favorable outcome among thinner than thicker primary tumors when controlling for number of pathologically involved nodes. This latter finding necessitated reclassification within the AJCC stage III (node positive) group.

Despite these apparent improvements in AJCC stage subclassification, a number of unanswered questions remain to be answered. How does patient ethnicity and tumor biology affect outcome, independent of disease stage and variability in treatment? Why do certain colon cancers (specifically node-negative) recur after apparently curative resection? What patients should be selected for more aggressive adjuvant systemic therapy, and which patients are less likely to respond to treatment and more likely to experience unnecessary treatment-related risks for little therapeutic gain? While the answers to some of these questions may remain unknown, multivariate statistical modeling may provide key insights into tumor behavior and provide survival estimates specific to an individual patient.

A machine-learned Bayesian belief network (ml-BBN) is a hierarchical network of associations between clinical factors in a registry data set that provides multivariate mapping of complex data. This is achieved through a straightforward, transparent graphical map. The ml-BBN model uses conditional dependence to provide a probability estimate of an outcome of interest. The model represents the dependence between independent variables associated with the outcome of interest. Importantly, it protects against overinterpretation of the data. Hence, the ml-BBN model encodes the joint probability distribution of all variables in a multidimensional data domain using a network of conditional probabilities.

The utility of Bayesian classification for developing clinical decision support (CDS) tools lies in the ability of the model to be queried. These model queries provide case-specific relative risk information for a given patient for an oncological outcome of interest. Clinical decision support systems (CDSS) based on ml-BBN models have been developed for a number of solid tumors in order to improve prognostic estimates and to guide clinical decision making for appropriate treatment and follow-up surveillance testing.[7–10] These CDSS have further assisted physicians in counseling of patients based on individualized estimates of prognosis.[11]

We sought to apply ml-BBNs to a large population in an effort to accurately derive individualized survival estimates using readily available case-specific clinical and pathological information. This study evaluated the feasibility of using ml-BBNs to provide real-time estimates of overall survival. These estimates were based on factors of prognostic significance for colon cancer in a retrospective analysis of data from the largest population-based cancer registry in the United States, the SEER Program of the National Cancer Institute (NCI). We found that the colon cancer ml-BBN developed in this study provides accurate time-dependent, individualized assessment of mortality risk.

METHODS

We used the SEER program of the National Cancer Institute (NCI), which comprises 17 registries across the United States and contains all patients diagnosed with

malignant colon cancer from 1969 to 2006. Of the original 437,892 records, a subset of 146,248 cases was selected for analysis based on these criteria: (1) diagnosis data subsequent to 2000 and (2) records complete. Staging was based on the American Joint Commission on Cancer (AJCC) sixth edition (after 2004) or third edition (between 2000 and 2004) TNM system.[12–15] TNM data were mapped to a simplified set of criteria consisting of M stage 0 or 1, N stage 0, 1, or 2, and T stage 1, 2, 3, or 4 or mapped to null when values could not be clearly mapped from AJCC third edition. As a result of our study set selection criteria, no variable had greater than 20 % missing data. SEER values for missing or unknown were all mapped to null and a data imputation method was applied. The imputation method used was a passive method specific to Bayesian learning referred to as "truncation" in which missing data were truncated from training records, and those data that were present used for learning structure and distributions.[16] Follow-up time and overall survival (OS) time were used to develop survival cohorts for ml-BBN modeling. To evaluate OS at different clinically relevant time points, 4 subsets were created. These subsets were based on follow-up time of 12, 24, 36, and 60 months. Cases within our study set were included in a follow-up cohort if they met 1 of 2 criteria: survival time in excess of the cohort window, or time to mortality less than that of the cohort window.

Statistical and Modeling Methods

Basic descriptive analysis of the study cohorts was performed using R:A Language and Environment for Statistical Computing.[17] Each cohort (12, 24, 36, and 60 months) was analyzed and clinical and pathological factors tabulated. Differences in distributions between subjects alive and dead within each follow-up cohort were analyzed with the *t* test (continuous variables), or the chi-squared test (categorical variables).

We used a stepwise modeling process as previously described.[9,10] Briefly, the stepwise process consists of: (1) preliminary modeling, (2) global modeling, and (3) focused or final modeling. This stepwise process has been used previously to produce effective, validated CDSS models.[10,11,18,19]

BBNs were created using machine-learning software (FasterAnalytics, DecisionQ Corp., Washington, DC). The machine-learning program uses heuristic algorithms to allow the computer to learn domain structure natively from data.[20–22] The software applies machine-learning algorithms to training data sets to learn BBN structures. BBNs are directed acyclic graphs of conditional dependencies between variables that allow users to understand how different variables interact. The ml-BBN allows calculation of posterior probability of an outcome (overall survival) given prior knowledge (e.g., clinical or pathological variables).[22,24] Our ml-BBN focuses on posterior estimates of mortality given known demographic, staging, pathological, treatment, and biomarker information.

A tenfold cross-validation was performed, and receiver operating characteristic (ROC) curves developed for each model. The data was randomized into 10 training sets containing 90 % of the data with 10 corresponding test sets containing the remaining 10 %. New ml-BBN models were created with the training sets and tested with the corresponding test set. An ROC curve was created for each test set and the area under the ROC curve (AUC) and 95 % confidence intervals (95 % CIs) calculated, creating a measure of how well the model could classify OS of an individual patient. To further validate the models, we also performed Kaplan–Meier analysis using the estimated mortality probabilities produced by the BBN as an additional validation method.

RESULTS

Using our inclusion criteria resulted in these study cohorts: the 12-month set contains 128,324 cases, the 24-month set contains 114,040 cases, the 36-month set 100,733 cases, and the 60-month set contains 77,402 cases. Overall mortality for each study cohort was 25 % at 12 months, 38 % at 24 months, 50 % at 36 months, and 73 % at 60 months, respectively. Descriptive statistics are described in Table 1a, b. The tabulated data is stratified by study cohort.

Table 2a, b provides comparative distributions by study cohort (12, 24, 36, and 60 months) between subjects who were deceased and living at the selected follow-up time for each cohort. The differences in distribution of the deceased and the living group for each covariate are statistically significant ($p < .001$).

A tenfold cross validation was used to assess ml-BBN model robustness in estimating predictive value for mortality. Each study cohort model, as cross-validated, is strongly predictive. The AUCs for each of the 12-, 24-, 36-, and 60-month cohorts are 0.85 with 95 % CIs of (0.84, 0.87), (0.84, 0.86), (0.84, 0.87), and (0.84, 0.87), respectively. Positive predictive values for mortality at 12, 24, 36, and 60 months are, 74, 80, 82, and 84 %, respectively. Negative predictive values for mortality at 12, 24, 36, and 60 months are, 85, 80, 74, and 65 %, respectively. Sensitivity (95 % CIs) for 12, 24, 36, and 60 months are 51 % (51–52 %), 63 % (61–64 %), 70 % (67–72 %), and 89 % (88–89 %), respectively. Specificity (95 % CIs) for the 12, 24, 36, and 60 month cohorts are 94 % (94–94 %), 90 % (90–91 %), 85 % (83–86 %), and 56 % (54–57 %), respectively. PPV, NPV, sensitivity, and specificity were evaluated using the "most likely" estimate for mortality

TABLE 1 Patient characteristics by cohort

A.

	12-month cohort (%) *N* = 128,324	24-month cohort (%) *N* = 114,040	36-month cohort (%) *N* = 100,733	60-month cohort (%) *N* = 77,402
Sex				
Female	52.2	52.3	52.5	52.8
Male	47.8	47.7	47.5	47.2
Tumor grade				
1	8.7	8.5	8.1	7.3
2	57.1	56.4	55.5	52.9
3	17.9	18.2	18.6	20.1
4	0.9	0.9	0.9	1.0
AJCC TNM path T				
1	12.5	12.1	11.5	9.6
2	10.4	10.0	9.4	7.8
3	44.3	43.6	42.6	39.7
4	20.0	21.1	22.7	27.7
AJCC TNM path N				
0	63.8	63.5	62.8	59.7
1	19.9	19.7	19.5	19.6
2	13.3	13.5	14.0	16.1
AJCC TNM path M				
0	78.1	77.1	77.7	69.1
1	19.3	20.5	22.3	28.0
Number of primaries				
1	89.7	89.4	89.2	89.2
2	9.3	9.6	9.7	9.7
3	0.9	0.9	1.0	1.0
4	0.1	0.1	0.1	0.1
Tumor histology				
Adenocarcinoma	95.2	94.9	94.6	93.7
Other	4.8	5.0	5.4	6.3
Tumor location				
Left colon	9.2	9.2	9.1	8.9
Multiple	6.7	7.2	7.9	9.6
Right colon	46.3	46.1	46.2	46.3
Sigmoid colon	29.6	29.3	28.7	27.1
Transverse colon	8.3	8.2	8.2	8.2
Race/ethnicity				
Hispanic	8.3	8.0	7.8	7.5
White	81.8	82.0	82.0	82.2
Black	10.8	10.9	11.0	11.4
Asian	4.0	3.9	3.8	3.6
Other	3.4	3.3	3.1	2.9
Deaths				
Yes	25.2	38.3	49.8	72.8
No	74.8	61.8	50.2	27.2
Surgical procedure				
None	11.8	12.9	14.2	17.9
Other	5.3	5.3	5.2	4.5

TABLE 1 continued

A.

	12-month cohort (%) N = 128,324	24-month cohort (%) N = 114,040	36-month cohort (%) N = 100,733	60-month cohort (%) N = 77,402
Palliative	0.3	0.2	0.2	0.2
Partial colectomy	77.9	76.6	75.1	71.4
Total colectomy	3.0	3.1	3.2	3.3

B.

12-month cohort (%)		24-month cohort (%)		36-month cohort (%)		60-month cohort (%)	
N = 128,324		N = 114,040		N = 100,733		N = 77,402	
Range	%	Range	%	Range	%	Range	%
Regional nodes examined							
≤2	24.9	≤2	26.1	≤1	24.6	0	27.2
3–9	26.1	3–9	26.5	2–8	24.6	1–8	25.1
10–16	25.7	10–16	25.0	9–14	26.4	9–14	22.2
17+	20.9	17+	19.8	15+	21.6	15+	21.9
Regional nodes positive							
0	64.5	0	64.2	0	63.5	0	60.6
1	9.0	1	8.9	1	8.7	1–2	14.4
2–3	10.1	2–3	10.1	2–3	10.1	3–6	12.3
4+	13.3	4+	13.5	4+	14.0	7+	8.1
Age at diagnosis (years)							
≤61	27.3	≤61	26.4	≤60	25.3	≤63	26.3
62–72	25.5	62–72	25.3	61–71	22.2	64–74	26.0
73–80	23.7	73–80	24.1	72–80	27.1	75–82	25.1
81+	23.5	81+	24.2	81+	25.4	83+	22.6

(>50 %) to segment cases into high- and low-risk cohorts. This threshold was selected to provide a comparable metric across all models and to minimize the trade-off between sensitivity and specificity. Further, the difference in survival between the high-risk and low-risk cohorts were significant ($p < .001$) using the log-rank test. These significant differences in OS between low- and high-risk cohorts are reflected in the odds ratios for mortality: 17.1, 16.3, 13.9, and 8.8 for 12-, 24-, 36-, and 60-month cohorts, respectively.

We evaluated first-degree associates (independently predictive variables in the ml-BBN models) of mortality at 12, 24, 36, and 60 months. These relationships are important because they show variables that can be used to calculate mortality at each time point on a per-patient basis (individualized CDSS). The final ml-BBN model structures are shown in Figs. 1, 2, 3 and 4. Each box in the figure represents a feature within our study data set, while the arcs (edges) represent patterns of conditional dependence between features and allow one to understand which combinations of features provide the posterior estimate of mortality and can be read in conjunction with Table 3.

This colon cancer ml-BBN model structure shows how different variables in the model associate with one another to calculate estimates for OS. A key finding in this study is seen in Table 3, which shows time-dependent influence of key variables that form first-degree associations with oncological outcome (OS). Specifically, Table 3 describes those ml-BBN model variables that can be used to estimate subject-specific mortality at different times following initial treatment.

The ml-BBN provides clinically relevant input on those pieces of clinical and pathological information, which are most critical and which are most useful across multiple follow-up time periods. For example, the 12- and 24-month models have exactly the same structure. There were fewer first-degree independently predictive variable associations with mortality in the 36-month model than in the 12- or 24-month models, and even fewer first-degree associations in the 60-month model. As we look at longer survival times posttreatment for primary colon cancer, fewer factors ultimately influence long-term survival. However, as the first-degree variable associations differ, the ml-BBN models remain similar in overall structure.

TABLE 2 Deceased patient characteristics by cohort

A.

	12-month cohort (%)		24-month cohort (%)		36-month cohort (%)		60-month cohort (%)	
	N = 128,324		*N* = 114,040		*N* = 100,733		*N* = 77,402	
	Dead	Alive	Dead	Alive	Dead	Alive	Dead	Alive
Gender								
Female	54.5	51.5	53.9	51.2	53.5	51.5	53.2	51.8
Male	45.5	48.5	46.1	48.7	46.5	48.5	46.8	48.2
Tumor grade								
1	4.7	10.0	4.9	10.7	5.3	10.8	5.6	11.7
2	40.3	62.7	44.5	63.7	46.9	64.0	48.8	63.7
3	23.2	16.2	23.5	14.9	23.0	14.3	22.3	14.1
4	1.6	0.7	1.5	0.6	1.4	0.5	1.3	0.4
AJCC TNM Path T								
1	5.7	14.8	5.8	16.0	5.9	17.0	6.3	18.2
2	3.7	12.7	4.1	13.7	4.6	14.3	5.2	14.8
3	26.7	50.2	30.3	51.8	32.6	52.4	35.0	52.4
4	40.7	13.0	40.0	9.4	38.5	7.0	36.1	5.0
AJCC TNM Path N								
0	56.0	66.4	53.7	69.6	53.5	72.1	54.2	74.4
1	16.5	21.1	18.7	20.3	19.6	19.4	20.1	18.4
2	18.7	11.5	20.5	9.1	20.5	7.6	19.8	6.2
AJCC TNM Path M								
0	48.6	88.1	51.9	92.7	59.3	95.9	58.4	97.7
1	45.3	10.5	43.1	6.5	40.7	4.1	37.6	2.3
Number of primaries								
1	93.8	88.4	92.5	87.5	91.5	86.9	90.3	86.1
2	5.7	10.5	6.9	11.2	7.8	11.6	8.8	12.2
3	0.5	1.1	0.6	1.2	0.7	1.3	0.8	1.4
4+	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.2
Tumor histology								
Adenocarcinoma	88.5	97.5	90.6	97.6	91.5	97.6	92.2	97.8
Other	10.2	1.2	8.2	1.0	8.5	2.4	6.7	0.9
Tumor location								
Left	8.1	9.6	8.4	9.6	8.5	9.6	8.7	9.5
Multiple	17.7	2.9	14.6	2.6	13.3	2.4	12.3	2.2
Right	45.0	46.8	46.3	45.9	46.5	45.8	46.6	45.5
Sigmoid	21.1	32.4	22.5	33.5	23.4	33.8	24.3	34.5
Transverse	8.1	8.3	8.2	8.3	8.2	8.3	8.1	8.3
Race/ethnicity								
Hispanic	7.4	8.6	7.5	8.4	7.5	8.1	7.5	7.7
White	81.8	81.9	81.4	82.3	81.5	82.6	81.6	83.6
Black	12.7	10.2	12.6	9.9	12.5	9.6	12.3	9.0
Asian	2.9	4.3	3.2	4.3	3.3	4.3	3.3	4.2

It is noteworthy that from the AJCC TNM staging system, T- and M-stage covariates are both first-degree variables in all the survival windows. Nodal status is significant in all survival windows; however, the number of positive nodes is the first-degree associate in all models tested except for the 60-month model; in this late time period model number of positive nodes is supplanted by AJCC TNM N stage. In all the ml-BBN models, number of

TABLE 2 continued

B.

12-month cohort (%)			24-month cohort (%)			36-month cohort (%)			60-month cohort (%)		
N = 128,324			*N* = 114,040			*N* = 100,733			*N* = 77,402		
Range	% dead	% alive	Range	% dead	% alive	Range	% dead	% alive	Range	% dead	% alive
Regional nodes examined											
≤2	43.3	18.7	≤2	38.7	18.3	≤1	34.0	15.2	0	31.5	15.7
3–9	20.0	28.2	3–9	21.8	29.3	2–8	21.2	27.9	1–8	23.0	30.4
10–16	17.5	28.4	10–16	19.8	28.2	9–14	22.8	30.1	9–14	23.3	19.2
17+	11.7	24.0	17+	13.7	23.6	15+	16.7	26.3	15+	17.3	34.3
Regional nodes positive											
0	57.7	66.8	0	55.2	69.8	0	54.8	72.0	0	55.4	74.4
1	6.1	10.0	1	7.0	10.0	1	7.5	9.9	1–2	14.2	15.2
2–3	8.6	10.6	2–3	10.1	10.1	2–3	10.7	9.6	3–6	14.0	7.5
4+	18.7	11.5	4+	20.5	9.1	4+	20.5	7.6	7+	12.7	0.0
Age at diagnosis (years)											
≤61	15.7	31.2	≤61	18.4	31.3	≤60	17.8	32.7	≤63	22.5	36.5
62–72	19.5	27.6	62–72	20.4	28.3	61–71	19.8	24.6	63–74	23.7	32.2
73–80	25.0	23.3	73–80	24.6	23.8	72–80	27.0	27.2	75–82	26.3	21.9
81+	39.9	17.9	81+	36.6	16.6	81+	35.3	15.5	83+	27.5	9.4

positive nodes and AJCC TNM N stage are associated with one another. Importantly, unlike logistic regression models the absence of a first-degree associated variable does not necessarily render the ml-BBN model unusable, as other model features recursively estimate survival even when lacking certain case-specific variables.

DISCUSSION

Machine-learned BBNs present certain unique benefits pertaining to survival modeling of registry cohorts. First is their ability to represent complex, nonlinear relationships in a straightforward, graphical manner. Second, ml-BBNs support patient-specific survival estimates. Third, these CDSS provide the clinician and patient with a user-friendly interface. Prior research has shown machine-learned BBNs to be robust.[8,10,11,21,23] Furthermore, ml-BBNs have been shown to be an efficient way to represent complex information.[9,20,24] Finally, available CDSS literature emphasizes the importance of intuitive, user-friendly systems that provide easy access to information within existing clinical workflow.[12,24]

Our results show that the methodology and approach we applied to the SEER registry data can produce robust predictive models of mortality in colon cancer using ml-BBNs, and that these models have the potential to dramatically improve individualized estimation of prognosis. In addition to providing subject-specific estimates (as opposed to cohort-specific), our models achieved results that were equivalent, and in many cases superior to AJCC staging system alone. Table 4 details these results, with the ml-BBNs producing AUCs, sensitivity, and specificity within each estimation period, superior to AJCC-based estimates. Our probability threshold for stratifying subjects into high- and low-risk cohorts was designed to minimize the trade-off between sensitivity and specificity rather than to clinically optimize for rule-in/rule-out of patients relative to treatment. In a planned future clinical deployment, we will report on sensitivity and specificity at a range of predicted probabilities based on the ROC curve analysis in order to allow clinicians and patients to use results at their own level of risk tolerance. In this manner, the models are adaptable to both patient and physician needs.

When used as a tool to estimate probability of outcome, each ml-BBN model has the ability to provide a patient-specific estimate of survival at a specific follow-up time point (12, 24, 36, and 60 months). This personalized estimate of mortality uses readily obtainable information about patient demographics, clinical parameters, and disease staging. To illustrate this point, Table 5 details an example in which we evaluate a 63-year old black male whose colon cancer is AJCC Stage N0M0. Our first calculation shows the case assuming a T3 tumor, with an 11 % probability of mortality—whereas a T4 tumor increases this probability to 30 %. Within the T4 tumor sizing, a histologic grade 1 tumor is associated with a 22 % probability of mortality within 12 months, while a histologic grade 4 tumor increases this probability to 54 %. With these types of

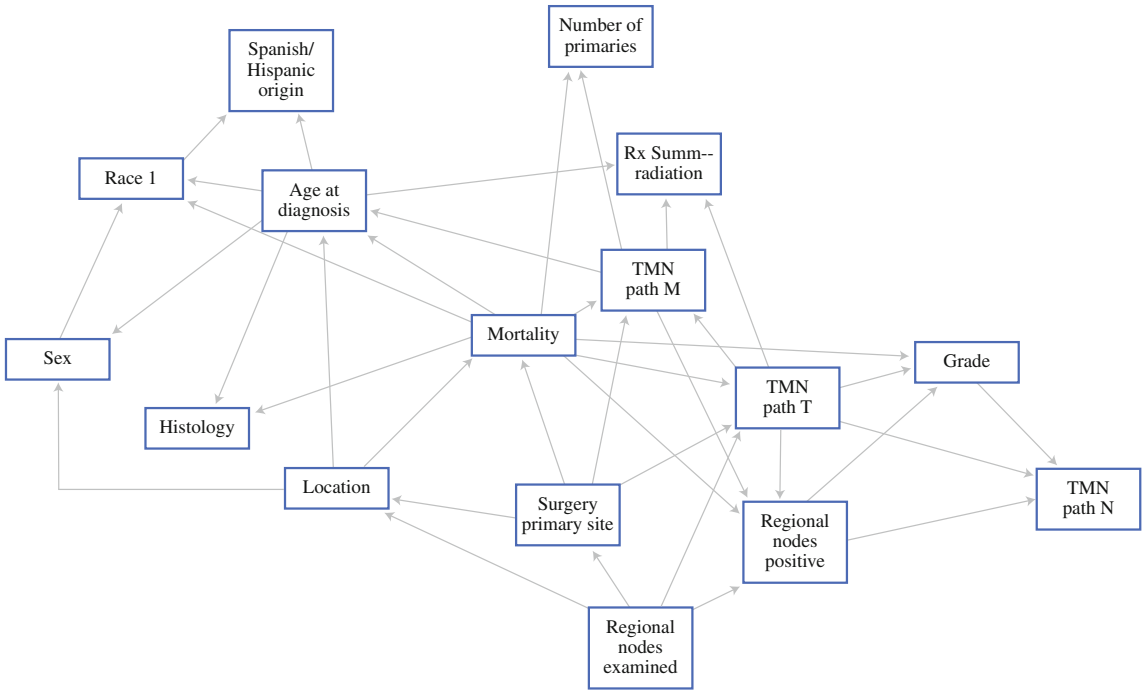

FIG. 1 12-month Bayesian belief network

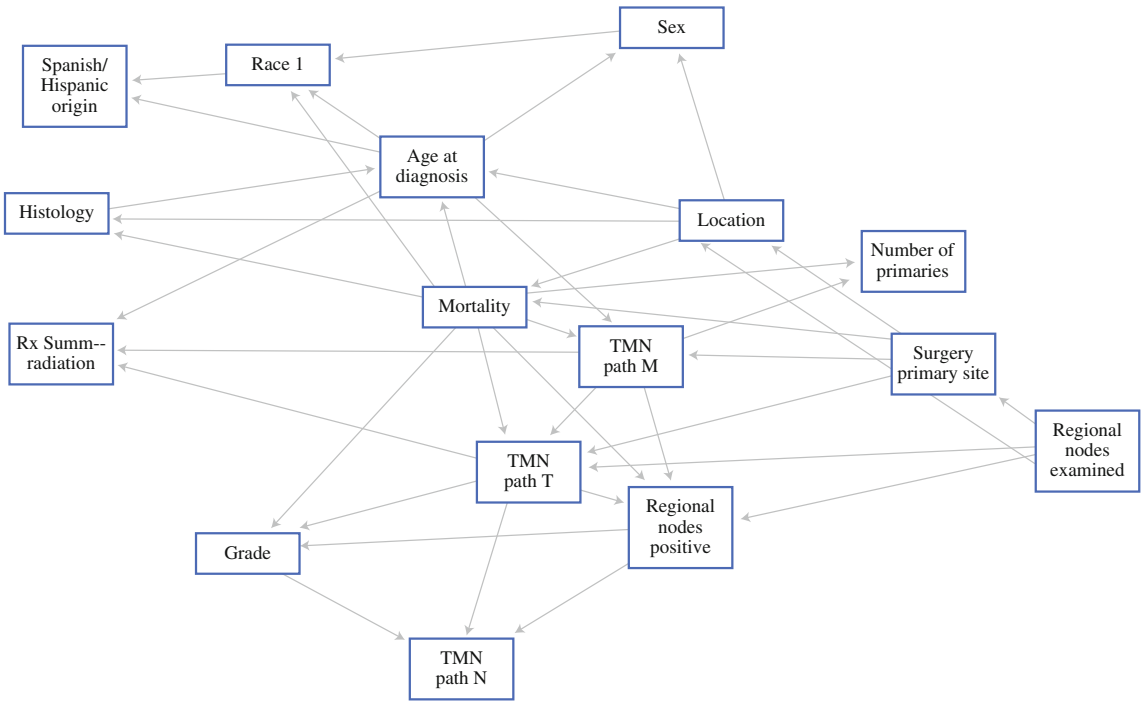

FIG. 2 24-month Bayesian belief network

case-specific estimates, the clinician and patient can work together to develop a therapy plan, which takes into account individual patient risk estimates as well as patient risk tolerance.

Recent literature has shown substantial variance in outcomes within AJCC TNM stages, emphasizing the need for system-based approaches that can develop personalized estimates of outcome using individual case

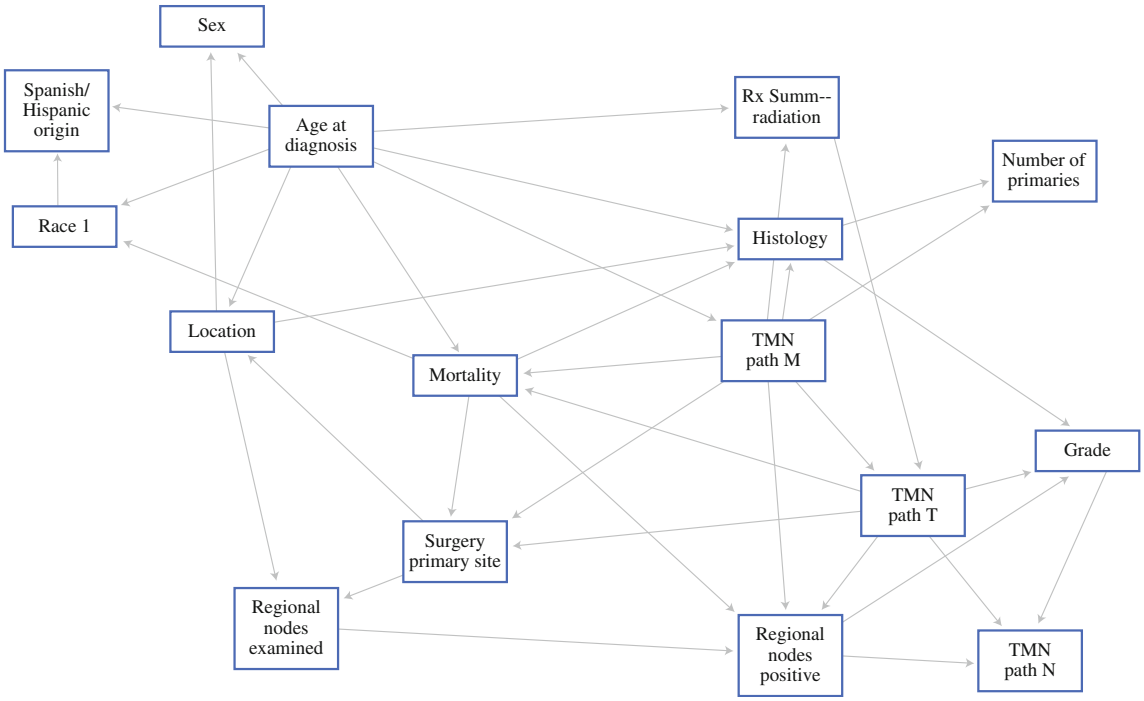

FIG. 3 36-month Bayesian belief network

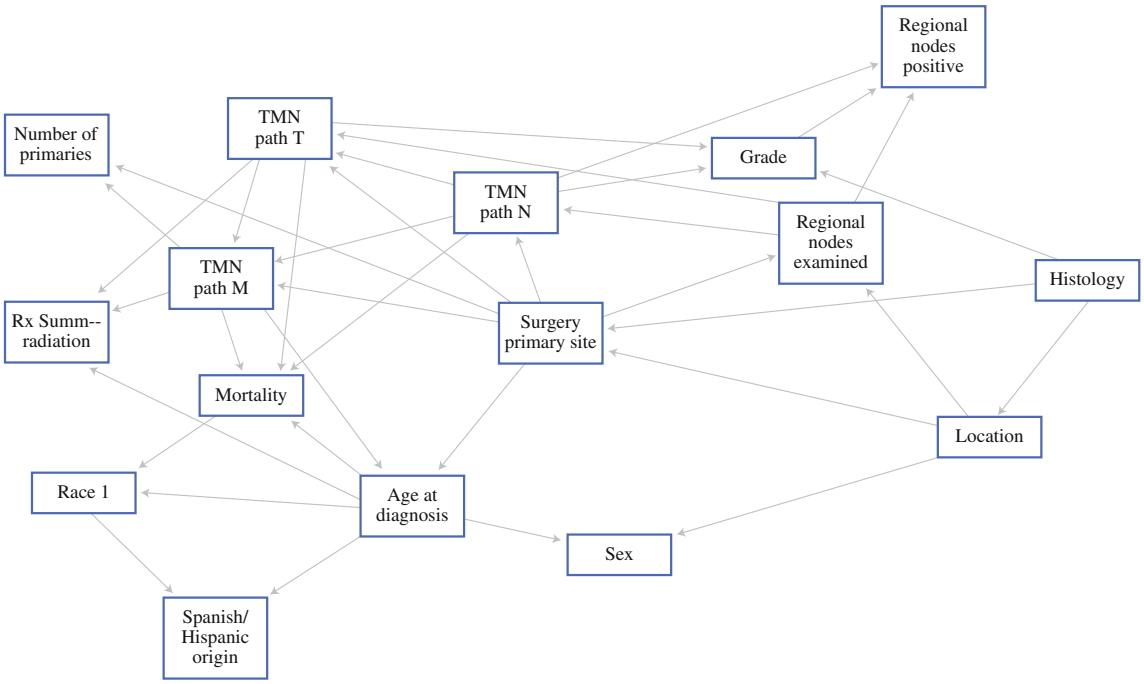

FIG. 4 60-month Bayesian belief network

attributes (specifically, readily available clinical and pathological variables).[25] The ml-BBN models provide information about how the disease factors influence outcome and how the disease factors influence each other. Further, these models extend AJCC TNM staging to incorporate information about patient demographics, tumor grading, and histology, in order to provide patient-specific assessments of risk that can in turn be used to make informed treatment decisions. In testing, these ml-BBN models have been shown to improve on the sensitivity and specificity of the AJCC TNM Staging (Sixth Edition) alone.

TABLE 3 First-degree variable associations to mortality in each ml-BBN mortality model

First degree associations to mortality	12-month model	24-month model	36-month model	60-month model
Number of primaries	Yes	Yes	No	No
AJCC TNM path M	Yes	Yes	Yes	Yes
Tumor grade	Yes	Yes	No	No
Race/ethnicity	Yes	Yes	Yes	Yes
Sex	Yes	Yes	Yes	No
Age at diagnosis	Yes	Yes	Yes	Yes
Regional nodes positive	Yes	Yes	Yes	No
Regional nodes examined	Yes	Yes	Yes	No
AJCC TNM path T	Yes	Yes	Yes	Yes
Surgical procedure	Yes	Yes	Yes	No
Primary tumor location	Yes	Yes	No	No
Primary tumor histology	Yes	Yes	Yes	No
AJCC TNM path N	No	No	No	Yes

In addition, using the models to derive patient-specific survival estimates, we can also observe interesting patterns of data associations that engender novel insights and hypotheses. For example, number of positive nodes, nodes examined, and AJCC TNM N stage are highly associated with one another and with mortality. What is unexpected, however, is that the estimated number of positive nodes is directly influenced by the nodes examined, and if too few nodes are examined, then subjects are more likely to be in the deceased group independent of nodal staging. This may indicate the importance of adequate nodal retrieval and its potential role as a quality indicator of case-specific N staging.[26,27] Based on our ml-BBN model structure and distributions we know that AJCC TNM staging remains a useful predictor of mortality.

The ml-BBN model can also be used to derive charts of individual survival estimates, as in Table 6, which shows that the AJCC TNM M stage is an extremely significant predictor of outcome, while T and N stage differentiate between outcomes within M stage group. Tumor grade is an important independent predictive variable in the 12-month model, as shown in Table 7, but is eclipsed by the importance of AJCC T stage in the 60-month model analysis. The model structures also show this relationship. Grade is a first-degree associate of mortality in the 12-month model, but not in the 60-month model. These tables show how individual staging components influence survival at different times in the post-treatment follow-up period. AJCC T and N stage influence survival estimates in non-metastatic patients at both 12 and 60 months, as they provide additional information on the severity of non-metastatic disease. However, in patients with distant metastatic disease, T and N stage only influence survival estimates in the 12-month period, and they have a negligible effect on 60-month survival estimates. This may indicate that systemic disease may have lower short-term mortality in the context of less-aggressive local disease.

The colon cancer ml-BBN developed in this study has shown several key advantages. First, this technology allowed us to capture complex, nonlinear, and in some cases, nonobvious patterns in a very large and heterogeneous data set. The use of machine-learning technology has allowed us to thoroughly mine a large and complex registry set relatively quickly and to identify many different patterns that exist in the data that inform us about tumor biology. Second, ml-BBNs are graphical models, more easily allowing the user to interpret model structure and design in a straightforward, intuitive manner. Further, as the ml-BBNs are multi-dimensional constructs, the absence of any given feature or features does not necessarily render the model unusable. More importantly, other features in the model can be used to estimate recursively outcome even when certain data elements (clinical and pathological variables) are missing. This construct also makes the models somewhat resistant to data errors, as not only are independent features used to estimate outcomes, but are also used to estimate one another. Finally, ml-BBNs are computationally efficient CDSS and can be used to render many complex rule sets in a computationally efficient manner.

The value of our findings is threefold: (1) the methodological proof-of-concept that ml-BBNs can be applied to this type of registry data and produce very robust results

TABLE 4 Comparative performance statistics—AJCC TNM Staging (Sixth Edition) vs. ml-BBN

Mortality	AUC		PPV		NPV		Sensitivity		Specificity	
	AJCC	BBN	AJCC	BBN	AJCC	BBN	AJCC	BBN	AJCC	BBN
12 months	0.75	0.85	36.2 %	74.4 %	88.7 %	85.1 %	36.2 %	51.4 %	88.6 %	94.0 %
24 months	0.76	0.85	54.6 %	79.9 %	81.1 %	79.7 %	54.6 %	62.7 %	81.1 %	90.3 %
36 months	0.77	0.85	67.7 %	81.8 %	72.2 %	73.9 %	67.7 %	69.9 %	72.2 %	84.5 %
60 months	0.77	0.85	85.9 %	84.2 %	47.7 %	64.8 %	85.9 %	88.7 %	47.7 %	55.5 %

TABLE 5 Estimating 12-month survival given T, N, M stages, grade, age, and race

Independent prognostic factors in the ml-BBN model							Outcome	
Age at diagnosis	Tumor grade	Race	Regional nodes positive	Sex	TNM path M	TNM path T	Mortality	
							No	Yes
61–72		AA	Up to 0	M	0	3	88.9	11.1
61–72		AA	Up to 0	M	0	4	69.6	30.4
61–72	1	AA	Up to 0	M	0	4	78.2	21.8
61–72	4	AA	Up to 0	M	0	4	45.7	54.3

TABLE 6 12- and 60-month mortality according to AJCC (v6) TNM stage

TNM path T	TNM path N	TNM path M	Alive at 12 months		TNM path T	TNM path N	TNM path M	Alive at 60 months	
			Yes	No				Yes	No
T1	N0	M0	88.5	11.5	T1	N0	M0	51.3	48.7
T2	N0	M0	91.4	8.6	T2	N0	M0	55.4	44.6
T3	N0	M0	88.2	11.8	T3	N0	M0	47.4	52.6
T4	N0	M0	67.8	32.2	T4	N0	M0	15.8	84.2
T1	N1	M0	90.9	9.1	T1	N1	M0	46.6	53.4
T2	N1	M0	91.4	8.6	T2	N1	M0	48.1	51.9
T3	N1	M0	86.2	13.8	T3	N1	M0	35.9	64.1
T4	N1	M0	70.6	29.4	T4	N1	M0	11.5	88.5
T1	N2	M0	78.4	21.6	T1	N2	M0	41.5	58.5
T2	N2	M0	88.0	12.0	T2	N2	M0	42.4	57.6
T3	N2	M0	78.2	21.8	T3	N2	M0	19.2	80.8
T4	N2	M0	58.8	41.2	T4	N2	M0	5.8	94.2
T1	N0	M1	26.6	73.4	T1	N0	M1	1.5	98.5
T2	N0	M1	48.8	51.2	T2	N0	M1	8.8	91.2
T3	N0	M1	52.4	47.6	T3	N0	M1	0.9	99.1
T4	N0	M1	29.3	70.7	T4	N0	M1	3.0	97.0
T1	N1	M1	48.5	51.5	T1	N1	M1	8.0	92.0
T2	N1	M1	61.4	38.6	T2	N1	M1	10.5	89.5
T3	N1	M1	62.8	37.2	T3	N1	M1	0.5	99.5
T4	N1	M1	55.4	44.6	T4	N1	M1	3.9	96.1
T1	N2	M1	52.1	47.9	T1	N2	M1	32.5	67.5
T2	N2	M1	50.4	49.6	T2	N2	M1	15.3	84.7
T3	N2	M1	48.6	51.4	T3	N2	M1	0.4	99.6
T4	N2	M1	43.5	56.5	T4	N2	M1	2.2	97.8

that are statistically comparable and in some instances superior to the AJCC staging system, (2) the insights that the ml-BBN graphical models provide about disease process and survival, and (3) the ability to use the inferential capability of the statistically validated ml-BBN models to calculate patient-specific estimates of survival. This is evidenced by the ROC AUCs, as well as the positive and negative predictive values reported previously. The SEER data, coupled with our ml-BBN model formulation and classifier was able to predict accurately patient outcomes given sufficient model flexibility and interdependence among prognostic factors. We expect these results to be reproducible for other disease sites outside of colon cancer. Given highly varied outcomes, the ability to use advanced computer modeling approaches not only to better understand disease outcomes but also to understand how they apply to a specific patient will be important in patient care and clinical trial design as well as incorporation of CDSS in the clinical pathway. The high AUCs (0.85) of our models show that the ml-BBNs have a high discriminatory capacity in estimating survival within a defined period following initial cancer treatment. These high AUCs are

TABLE 7 12- and 60-month mortality with AJCC (v6) TNM T stage and primary tumor grade as primary ml-BBN drivers (independent prognostic factors in the ml-BBN model)

Tumor grade (differentiation)	TNM Path T	Alive at 12 months		Tumor grade (differentiation)	TNM Path T	Alive at 60 months	
		Yes	No			Yes	No
Well differentiated	T1	89.4	10.6	Well differentiated	T1	48.0	52.0
	T2	91.7	8.3		T2	53.6	46.4
	T3	87.1	12.9		T3	40.0	60.0
	T4	55.0	45.0		T4	4.8	95.2
Moderately differentiated	T1	86.9	13.1	Moderately differentiated	T1	47.8	52.2
	T2	91.0	9.0		T2	53.1	46.9
	T3	86.3	13.7		T3	37.8	62.2
	T4	50.3	49.7		T4	4.9	95.1
Poorly differentiated	T1	75.1	24.9	Poorly differentiated	T1	45.7	54.3
	T2	88.9	11.1		T2	52.0	48.0
	T3	77.5	22.5		T3	32.6	67.4
	T4	36.4	63.6		T4	4.5	95.5
Undifferentiated	T1	62.5	37.5	Undifferentiated	T1	38.9	61.1
	T2	84.3	15.7		T2	50.6	49.4
	T3	68.0	32.0		T3	31.9	68.1
	T4	26.8	73.2		T4	4.2	95.8

further confirmed using Kaplan-Meier (log rank) analysis that shows high, statistically significant odds ratios. Finally, when compared with the AJCC staging system alone, our ml-BBN models showed superior sensitivity and specificity in estimating mortality. Our methodology has the potential to dramatically expand the toolsets available to clinicians in order to tailor personalized treatment plans, to improve outcomes, and to empower patients through education about their disease while leveraging robust CDSS.

While the models developed in this study do not explicitly provide an estimate of response to therapy, we believe that these recurrence and mortality risk assessment methods can be used at the point of care by physicians to estimate outcomes that are much more specific for an individual patient. This will then facilitate patient education and counseling. Two well-known examples currently in use are Adjuvant!Online and the Memorial Sloan Kettering Colon Cancer Nomogram. Adjuvant!Online was designed as an evidence-based tool to assist clinicians with evaluating the potential benefit of chemotherapy in breast cancer patients.[28] It uses the SEER database to derive a background estimate of mortality and then uses literature-based estimates for chemotherapy benefit.[29] A more relevant example is the Memorial Sloan Kettering Colon Cancer Nomogram, which includes chemotherapy, and uses a 1,320-subject institutional registry cohort, yielding a concordance index (a metric similar to an AUC) of 0.77.[30] Both the Memorial Sloan Kettering nomogram and our large SEER-derived ml-BBN model allow the clinician to input case-specific information and to derive patient-specific estimates of survival. Finally, a current example of how prognostic estimates are used in guiding therapy is the OncotypeDX test marketed by Genomic Health for use in node-negative, ER/PR positive breast cancer.[31] The OncotypeDX does not explicitly indicate the likely benefit that a patient will receive from chemotherapy, in fact 1 study shows that patients in all risk cohorts receive some benefit from chemotherapy.[32] While the data would appear to show that all patients benefit, not all patients have the same magnitude of benefit; hence prognostic tools allow clinicians to develop an estimate of the potential magnitude of therapeutic benefit for a specific patient.

Finally, we recognize certain limitations in this study. The SEER database population is predominantly Medicare/Medicaid based, and it tends to have a bias toward older subjects and among the older records, toward white subjects. Second, the SEER database lacks recurrence and chemotherapy data, which prohibits us from drawing direct inference about these important factors. The primary methodological limitation in computer modeling is the risk of overfitting the data—of trying to draw broad conclusions from a model that is representative of a study sample. To address this risk, we: (1) used a large sample size, (2) made use of machine-learning algorithms that are designed to produce robust models, and (3) performed cross-validation analysis to evaluate interset variance within the ml-BBN models. Narrow confidence intervals between cross-validation sets

imply that these models are robust, indeed. It is also important to recognize that the SEER database is not sensitive generally to the noteworthy improvements over time that have advanced treatment for this disease. Finally, confounding the issue is the lack of any data on chemotherapy and limited data on radiotherapy within the SEER database.

In summary, we were able to successfully apply a computer-based CDSS tool using ml-BBNs to provide personalized estimates of survival in colon cancer. These ml-BBNs were based on patient demographics and clinical and staging information at specific post-treatment follow-up times. Unlike traditional AJCC TNM staging, the colon cancer ml-BBN developed in this study accounts for tumor heterogeneity and missing information, and provides insights as to how multiple readily available patient and disease-specific variables collectively influence oncological outcome. As such, the colon cancer ml-BBN provides time-dependent, individualized mortality risk assessment to facilitate physicians making informed treatment decisions, thereby facilitating individual patient education about their disease and providing case-specific treatment recommendations. Further, this ml-BBN may ultimately improve clinical trial design.

ACKNOWLEDGEMENT This clinical research effort was supported, in part, by the United States Military Cancer Institute, Washington, D.C. and the Henry M. Jackson Foundation, Rockville, MD and RO1 Grant CA090848.

REFERENCES

1. Edwards BK, Ward E, Kohler BA, Eheman C, Zauber AG, Anderson RN, et al. Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. Cancer. 2010;116:544–73.
2. Vlug MS, Wind J, Hollmann MW, Ubbink DT, Cense HA, Engel AF, et al. Laparoscopy in combination with fast track multimodal management is the best perioperative strategy in patients undergoing colonic surgery: a randomized clinical trial (LAFA-study). Ann Surg. 2011;254:868–75.
3. Baldwin LM, Dobie SA, Billingsley K, Cai Y, Wright GE, Dominitz JA, et al., Explaining black-white differences in receipt of recommended colon cancer treatment, J Natl Cancer Inst. 2005;97:1211–20.
4. Cooper GS, Yuan Z, Landefeld CS. Surgery for colorectal cancer: Race-related differences in rates and survival among Medicare beneficiaries. *Am J Public Health.* 1996;86:582–6.
5. Jessup JM, Gunderson LL, Greene FL, Washington MK, Compton CC, Sobin LH, et al. 2010 staging system for colon and rectal carcinoma. *Ann Surg Oncol.* 2011;18:1513–7.
6. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, eds. *AJCC Cancer Staging Manual.* 7th ed. New York: Springer, 2010.
7. Burnside ES, Rubin DL, Fine JP, Shachter RD, Sisney GA, Leung WK. Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology.* 2006;240:666–73.
8. Maskery SM, Hu H, Hooke J, Shriver CD, Liebman MN. A Bayesian derived network of breast pathology co-occurrence. J Biomed Inform. 2008;41:242–50.
9. Stojadinovic A, Eberhardt C, Henry L, Eberhardt JS, Elster EA, People GE, et al. Development of a Bayesian classifier for breast cancer risk stratification: a feasibility study. *Eplasty.* 2010; 10:e25.
10. Stojadinovic A, Peoples GE, Libutti SK, Henry LR, Eberhardt J, Howard RS, et al. Development of a clinical decision model for thyroid nodules. *BMC Surg.* 2009;9:12.
11. Berner ES. Clinical decision support systems: State of the Art. AHRQ Publication No. 09-0069-EF. Rockville, MD: Agency for Healthcare Research and Quality. June 2009.
12. Greene FL, Page DL, Fleming ID, Fritz A, Balch CM, Haller DG, et al., eds. *AJCC Cancer Staging Manual.* 6th ed. New York: Springer. 2002.
13. Classification for Extent of Disease, Self Instructional Manual for Tumor Registrars, Surveillance, Epidemiology, End Results Program, National Cancer Institute, National Institute of Health; Bethesda, MD, 1977.
14. SEER Summary Stage Manual, 2000, Surveillance, Epidemiology, End Results Program, National Cancer Institute, National Institute of Health; Bethesda, MD, 2001.
15. SEER Extent of Disease (EOD), Surveillance, Epidemiology, End Results Program, National Cancer Institute, National Institute of Health; Bethesda, MD, 1998.
16. Lavoti PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. *Stat Med.* 1995; 14:1913–25.
17. *R: A Language and Environment for Statistical Computing, R Development Core Team.* Vienna, Austria: R Foundation for Statistical Computing. 2011. ISBN 3-900051-07-0, http://www.R-project.org.
18. Forsberg JA, Eberhardt J, Boland PJ, Wedin R, Healey JH. Estimating survival in patients with operable skeletal metastases: an application of a Bayesian belief network. *PLoS One.* 2011;6: e19956.
19. Stojadinovic A, Nissan A, Eberhardt J, Chua TC, Pelz JO, Esquivel J. Development of a Bayesian belief network model for personalized prognostic risk assessment in colon carcinomatosis. *Am Surg.* 2011;77:221–30.
20. Jensen F. *An Introduction to Bayesian Networks.* New York: Springer-Verlag, 1996.
21. Moraleda J, Miller T. Ad+tree: A compact adaptation of dynamic ad-trees for efficient machine learning on large data sets. *Proceedings of the 4th International Conference on Intelligent Data Engineering and Automated Learning.* 2002.
22. Pearl J. (1988) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA: Morgan Kaufmann, San Mateo, CA, 1988.
23. Stojadinovic A, Eberhardt J, Brown TS, Hawksworth JS, Gage F, Tadaki DK, et al. Development of a Bayesian model to estimate health care outcomes in the severely wounded. J Multidiscip Healthc. 2010;3:125–35.
24. Osheroff JA, Teich JM, Middleton BF, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc. 2007;14:141–5.
25. Ueno H, Mochizuki H, Shirouzu K, Kusumi T, Yamada K, Ikegami M, et al. Actual status of distribution and prognostic impact of extramural discontinuous cancer spread in colorectal cancer. *J Clin Oncol.* 2011;29:2550–6.
26. Smith DD, Schwarz RR, Schwarz RE. Impact of total lymph node count on staging and survival after gastrectomy for gastric

cancer: Data from a large U.S. population database. *J Clin Oncol.* 2005;23:7114–24.
27. Schwarz RE, Smith DD. Extent of lymph node retrieval and pancreatic cancer survival: information from a large US population database. *Ann Surg Oncol.* 2006;13:1189–2006.
28. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol.* 2001;19:980–91.
29. Gribbin J, Dewis R; Adjuvant! Online: review of evidence concerning its validity, and other considerations relating to its use in the NHS; National Collaborating Centre for Cancer (UK). Early and Locally Advanced Breast Cancer: Diagnosis and Treatment [Internet]. Cardiff (UK): National Collaborating Centre for Cancer (UK); 2009 Feb. (NICE ClinicalGuidelines, No. 80).
30. Weiser MR, Landmann RG, Kattan MW, Gonen M, Shia J, Chou J, et al. Individualized prediction of colon cancer recurrence using a nomogram. *J Clin Oncol.* 2008;26:380–5.
31. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *N Engl J Med.* 2004;351:2817–26.
32. Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, Pho M, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res.* 2006;8:R25.