

Predicting Lung Cancer Survival Using Probabilistic Reclassification of TNM Editions With a Bayesian Network

Melle S. Sieswerda, MD, MSc^{1,2}; Inigo Bermejo, PhD²; Gijs Geleijnse, PhD¹; Mieke J. Aarts, PhD¹; Valery E.P.P. Lemmens, PhD²; Dirk De Ruyscher, MD, PhD²; André L.A.J. Dekker, PhD²; and Xander A.A.M Verbeek, PhD¹

PURPOSE The TNM classification system is used for prognosis, treatment, and research. Regular updates potentially break backward compatibility. Reclassification is not always possible, is labor intensive, or requires additional data. We developed a Bayesian network (BN) for reclassifying the 5th, 6th, and 7th editions of the TNM and predicting survival for non–small-cell lung cancer (NSCLC) without training data with known classifications in multiple editions.

METHODS Data were obtained from the Netherlands Cancer Registry (n = 146,084). A BN was designed with nodes for TNM edition and survival, and a group of nodes was designed for all TNM editions, with a group for edition 7 only. Before learning conditional probabilities, priors for relations between the groups were manually specified after analysis of changes between editions. For performance evaluation only, part of the 7th edition test data were manually reclassified. Performance was evaluated using sensitivity, specificity, and accuracy. Two-year survival was evaluated with the receiver operating characteristic area under the curve (AUC), and model calibration was visualized.

RESULTS Manual reclassification of 7th to 6th edition stage group as ground truth for testing was impossible in 5.6% of the patients. Predicting 6th edition stage grouping using 7th edition data and vice versa resulted in average accuracies, sensitivities, and specificities between 0.85 and 0.99. The AUC for 2-year survival was 0.81.

CONCLUSION We have successfully created a BN for reclassifying TNM stage grouping across TNM editions and predicting survival in NSCLC without knowing the true TNM classification in various editions in the training set. We suggest binary prediction of survival is less relevant than predicted probability and model calibration. For research, probabilities can be used for weighted reclassification.

JCO Clin Cancer Inform 4:436–443. © 2020 by American Society of Clinical Oncology

Licensed under the Creative Commons Attribution 4.0 License 

INTRODUCTION

In cancer care, the TNM system for classification of malignant tumors guides treatment decisions, aids in stratifying patients for research, and helps clinicians assess prognosis.^{1,2} This is done by classifying characteristics of the tumor (T descriptor), local lymph nodes (N descriptor), and distant metastases (M descriptor). These descriptors can subsequently be used to compute a stage grouping, essentially summarizing the information.

The system is revised on a regular basis (every 5.7 years on average), during which changes are made both to the individual descriptors and to the stage grouping.³ Revisions incorporate new developments that improve (outcome) stratification and prognostic capabilities, keeping the classification system relevant. However, because categories can be added or removed, classes with the same label are not necessarily

equivalent across editions.⁴ Recommendations for care interventions from the literature and clinical trials may be based on specific editions of the classification system, and it is not always immediately clear how to apply these recommendations to patients classified with a different edition.⁵ Also, scientific analysis of patient cohorts classified with different editions must consider the differences across editions.

This issue can be tackled by either mapping class labels from source to target edition or by (re)classifying the patient using the target edition. If mapping is not feasible, additional data are required to help determine the individual descriptors in the target edition. In practice, this process is complicated, and these data are usually excluded from the analyses or an approximate mapping is assumed.

For both situations described, it would be helpful to have a model that can aid in the reclassification across

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on January 31, 2020 and published at ascopubs.org/journal/cci on May 11, 2020; DOI <https://doi.org/10.1200/CCI.19.00136>

CONTEXT**Key Objective**

The TNM classification system is used for prognosis, treatment, and research; however, data classified with different editions are not directly comparable. Reclassifying data across TNM editions would facilitate pooling historical data for research.

Knowledge Generated

Bayesian networks can be used to reclassify stage grouping across TNM editions by combining clinical knowledge with real-world data and leveraging the relation between TNM stage and survival.

Relevance

Probabilistic reclassification across TNM editions for NSCLC makes it possible to use data classified with different TNM editions for research. Secondly, our Bayesian network offers multiclass survival prediction with accompanying probabilities, possibly helping to assess prognosis.

TNM editions. In this article, we develop such a model based on Bayesian networks (BNs).

A BN is a type of probabilistic graphic model that uses a directed acyclic graph. The nodes represent variables, and the directed edges signify (preferably causal) relationships.^{6,7} Each node is associated with a probability distribution that is conditional on its parents (ie, the set of nodes that have a directed edge to that node), which can be written as $P(X|Pa_X)$. This leads to a set of (conditional) probability distributions that together define a joint probability distribution, which can be written as $P(X_1, \dots, X_n) = \prod_{i=0}^n P(X_i|Pa_{X_i})$. For nodes that are associated with a discrete probability distribution, the distribution is defined as a conditional probability table (CPT).

BNs can be used to estimate the probability distribution of a variable given evidence. In contrast to other models, such as logistic regression, BNs do not have dedicated inputs or outputs. Instead, setting evidence on any node updates probabilities throughout the network.

The resulting models can easily be evaluated by medical specialists: the graphic nature of BNs makes interpretation of relationships straightforward, and conditional probability aligns well with physicians' reasoning. This is a benefit over black-box approaches, such as artificial neural networks or deep learning.

In this work, we hypothesized BNs can be used to reclassify data across TNM editions for non-small-cell lung cancer (NSCLC) and predict survival. We further hypothesized that such a BN can be learned without knowing the true TNM classification in various editions in the training set by leveraging the correlation between TNM and survival.

METHODS**Data**

Data were obtained from the population-based Netherlands Cancer Registry (NCR) after approval by the NCR Privacy Review Board and did not require approval from an ethics committee in the Netherlands. The NCR has been maintained since 1989 and is populated by trained data managers. It contains all cancer occurrences in the Netherlands. Coding is based on international rules and standards. The edition of the TNM classification used depends on the incidence year (Table 1).

Inclusion criteria were pathology confirmed, International Classification of Diseases O-3 topology code C34 (bronchus and lung), morphology codes appropriate for NSCLC, and incidence year between 1999 and 2016; 2017 and 2018 were excluded because of insufficient follow-up. Patients with multiple primary tumors were kept because lung cancer is generally dominant in determining survival. A total of 146,084 patients fulfilled our criteria. We obtained the following variables: clinically staged T, N, and M descriptors; incidence year; days of follow-up; and vital status at follow-up. A new variable was added to each record specifying the TNM edition used, based on year of incidence. Survival time, counted from initial diagnosis, was discretized into 5 categories frequently used in the literature and clinical practice (Table 2). Discretization introduced missing values for patients who had a follow-up of less than 2 years and were alive at the time of follow-up. The dataset was randomly split into a training set (80%; $n = 116,858$) and a test set (20%; $n = 29,226$). Because the distribution of the different TNM editions was unequal, the training set was resampled (with replacement) to contain 45,000 samples for each edition. NAs were removed from the test set.

Network Structure Definition

A BN was designed to predict the 5th and 6th TNM editions with variables from the 7th edition of the TNM classification

TABLE 1. TNM Editions Used by the Netherlands Cancer Registry and No. of Records Available by Period, Split Into Data for Training and Testing

Period	TNM Edition	No. (training)	No. (training, after resampling)	No. (test)
1999-2002	5	21,528	45,000	5,323
2003-2009	6	43,952	45,000	11,061
2010-2016	7	51,378	45,000	12,842

TABLE 2. Parameters Obtained From the Netherlands Cancer Registry

Parameter	Values	Description
T	T1	Clinical T descriptor (ie, cT). The values T0 and Tis were excluded because of the low frequency in the cancer registry data
	T1a	
	T1b	
	T2	
	T2a	
	T2b	
	T3	
	T4	
	TX	
N	N0	Clinical N descriptor (ie, cN)
	N1	
	N2	
	N3	
M	M0	Clinical M descriptor (ie, cM)
	M1	
	M1a	
	M1b	
TNM	1a	Clinical stage group (ie, cTNM)
	1b	
	2a	
	2b	
	3a	
	3b	
	4	
	X	
Incidence year	Integer	Year of incidence
Days of follow-up	Integer	Years of follow-up available
Vital status	Boolean	Vital status during last follow-up
Survival	< 30 days	Time of death after diagnosis. Calculated using days of follow-up and vital status
	1-4 months	
	4-6 months	
	6-12 months	
	1-2 years	
Multifocal	Boolean	True if the tumor was multifocal
Tumor size	Integer	Diameter in mm. Only available for patients with incidence year \geq 2015
Sulcus superior	Boolean	Indicates involvement of the sulcus superior. Only available for patients with incidence year \geq 2015

NOTE. The variables "Multifocal," "Tumor size," and "Sulcus superior" were only used for computing the test set and not for training the model.

(TNM7) and vice versa, shown in [Figure 1](#), using BayesiaLab 8.⁸ This required 2 sets of 4 nodes corresponding to T, N, and M descriptors and TNM stage grouping: one set for input and one for output. By adding 2 nodes for TNM edition and survival, a total of 10 nodes was obtained. Relationships between nodes were established to indicate causal effect. Nodes edition, T_567, N_567, M_567, TNM_567, and death were

associated with variables in the dataset, leaving T_7, N_7, M_7, and TNM_7 as hidden nodes. CPTs for the hidden nodes were estimated on a subset of the training set containing only 7th edition data. Relationships between hidden nodes and their observed counterparts were primed by manually estimating CPTs $P(T_{567}|edition, T_7)$, $P(N_{567}|edition, N_7)$, and $P(M_{567}|edition, M_7)$ through analysis of the

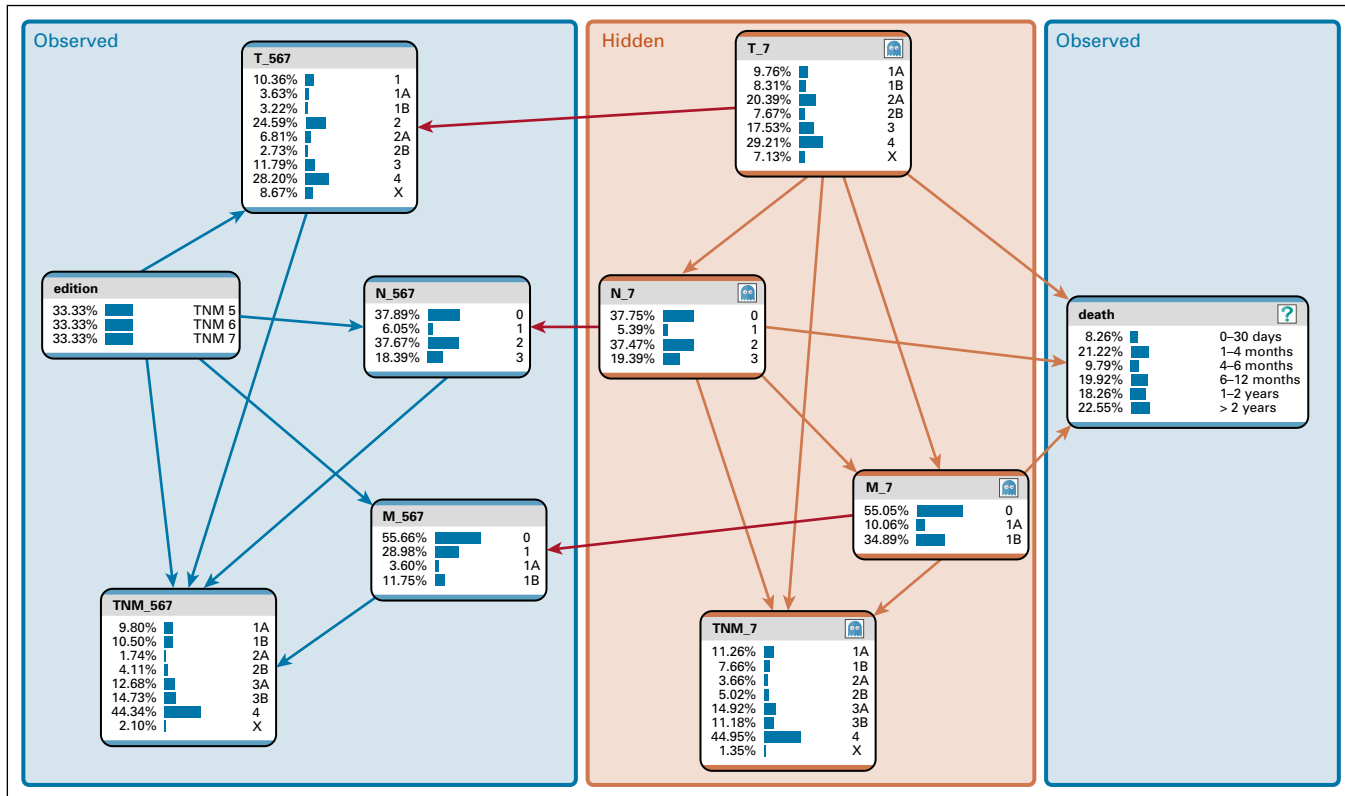


FIG 1. Bayesian network (after Expectation-Maximization learning) visualized using BayesiaLab 8. Each node represents a random variable with a (conditional) probability table and shows its state’s prior probabilities in a (rotated) histogram. The arrows indicate the (causal) relations between the nodes. The “_567”-suffix indicate nodes that, in conjunction with the “edition” node, can take on values from all TNM editions. The “_7”-suffixed nodes can take on 7th edition values only.

differences between editions 6 and 7 in the American Joint Committee on Cancer (AJCC) staging manual (Data Supplement); there were no changes between editions 5 and 6, and survival differed only marginally (Data Supplement). Finally, Expectation-Maximization learning was used to estimate the CPTs of the full network.⁹

Reclassifying TNM7 Data as Ground Truth for Testing

Evaluating network reclassification performance requires a ground truth. Therefore, a subset of the test set was manually reclassified as follows.

The additional parameters of multifocal disease, tumor size, and sulcus superior involvement were obtained from the NCR for patients in the dataset (Table 2). Because collection of tumor size and sulcus superior involvement started in 2015, the subset was limited to include patients diagnosed in 2015 and 2016 (n = 3,544).

By comparing definitions for the T, N, and M descriptors in the AJCC staging manuals, reclassification rules were defined (Data Supplement). If a one-to-one mapping was impossible, determining a range of values was attempted. With the individually reclassified descriptors, the TNM stage group was computed.

Evaluating Predictive Network Performance

The network’s performance was evaluated on (1) predicting the 6th edition stage grouping, (2) predicting the 7th edition stage grouping using the ground truth set, and (3) predicting survival using the test set. Predicting the 5th edition was not evaluated separately, because the 5th and 6th editions are equivalent. The HUGIN Analysis Wizard was used to compute the confusion matrix for each evaluation by selecting the state with the highest belief as predicted state.

Macro-averaging and micro-averaging are used in multi-class problems to combine multiple metrics into single values (Data Supplement).¹⁰ Micro-averaged and macro-averaged sensitivity, specificity, and accuracy were computed with the Python programming language.¹¹⁻¹⁴ We additionally used BayesiaLab to calculate the receiver operating statistic area under the curve (ROC-AUC) for 2-year survival.

Model calibration for survival was visualized using a bubble plot and curve. Briefly, each unique combination of inputs (ie, edition T_567, N_567, M_567, and TNM_567) defines a subpopulation. The plots show, for each subpopulation, the relation between a survival category’s predicted probability and its observed frequency in the dataset. The size of each

bubble indicates the population's size in the dataset. The calibration curve, computed using scikit-learn in Python, averages predicted and observed values by applying quantile-based binning with 1,000 samples per bin (21 bins).^{14,15}

RESULTS

Network Structure and Parameters

The final network structure is shown in [Figure 1](#). Each node represents a random variable and shows its state's prior probabilities in a histogram. The arrows indicate (causal) relations between the nodes. A node's underlying CPT is conditional on its parents.

We assumed that the newest available TNM edition approximates the true disease state best and thus has a causal relation with survival. This also means that because the TNM edition used for staging a patient should not (causally) influence survival, each of the nodes T_567, N_567, and M_567 has 2 parents: edition and its corresponding unobserved/hidden counterpart.

After the network structure was created, Expectation-Maximization learning was successfully applied. Percentages shown on each node show the prior probabilities for each state. In general, the model behaved as expected: an increase in tumor stage in any of the T, N, M, and TNM nodes corresponded to poorer survival. Additionally, the relations between the hidden and the observed TNM nodes were close to the priors set manually.

Manual Reclassification of the Dataset

Multifocality, tumor size, and sulcus superior involvement were not always available for all patients. For 150 records (4%), the T descriptor could not be determined because of these missing values. A total of 451 patients (13%) had multifocal T4 tumors according to the 7th edition for which it was not possible to determine the corresponding 6th edition T descriptor without additional information. Another 216 records (6%) could only be classified into a range (eg, T2-T3). As a result, for 817 records (23%), the T descriptor could not be completely reclassified (Data Supplement). The N descriptor could be reclassified in all patients. There were 85 multifocal T4 tumors (2%) that did not have distant metastases in the 7th edition, making it impossible to determine the M descriptor in TNM6.

Computation of the stage group does not always require the T or N descriptor. For example, if a patient has distant metastases (M1), the stage group will always be IV. Also, in patients without distant metastases (M0), if the lymph node metastases are classified as N2, any T2 or T3 will yield a stage group of IIIA.

Consequently, in 3,344 of 3,544 records (94.4%), it was possible to definitively determine the stage group. In the remaining 200 patients (5.6%), only a range or a set of stage groups could be identified. Fifty-two patients were either stage IB or IIB, 13 were stage IIB or IIIA, and 85 were stage IIIA or IV. In 20 patients, the stage could range between IA and IIIB. Finally, in 6 patients, the stage could range between IIA and IIIB (Data Supplement). Partial classifications were removed from the ground truth set.

Evaluating Predictive Network Performance

Predicting 6th edition stage grouping using 7th edition data resulted in an average accuracy of 0.99. Macrosensitivity and microsensitivity were 0.92 and 0.96, respectively. Macrospecificity and microspecificity were 0.99 and 0.99, respectively. Inspection of the confusion matrix showed 2 situations where misclassification was notable. Forty-one percent of the patients who were stage IIB were misclassified as stage IB. Twenty-one percent of the patients classified as stage IIIB were misclassified as stage IB, IIB, or IIIA.

Predicting 7th edition stage grouping using 6th edition data yielded an average accuracy of 0.99. Macrosensitivity and microsensitivity were 0.85 and 0.95, whereas macrospecificity and microspecificity were 0.99 and 0.99, respectively. Again, inspection of the confusion matrix showed 2 notable situations where misclassification was apparent. Stage IIA was misclassified as IA in 53% of patients, and stage IIB was misclassified as IIA or IIIA in 59% of patients. Macro-averaged and micro-averaged sensitivity, specificity, and accuracy for survival are listed in [Table 3](#).

The confusion matrices underlying these statistics can be found online (Data Supplement), together with the sensitivity, specificity, and ROC-AUC for predicting each stage group (Data Supplement) and the corresponding set of ROC-AUC curves (Data Supplement). The ROC-AUC for

TABLE 3. Macro-Aggregated and Micro-Aggregated Statistics for Predicting 6th Edition Stage Group Using 7th Edition Data, 7th Edition Stage Group Using 6th Edition Data, and Survival Using Data From All Editions

Sensitivity/Specificity/ Accuracy	Predicting TNM6		Predicting TNM7		Predicting Survival	
	Macro-Aggregated	Micro-Aggregated	Macro-Aggregated	Micro-Aggregated	Macro-Aggregated	Micro-Aggregated
Sensitivity	0.918	0.957	0.851	0.948	0.282	0.350
Specificity	0.994	0.994	0.993	0.993	0.862	0.870
Accuracy	0.989	0.989	0.987	0.987	0.783	0.783

Abbreviations: TNM6, TNM classification, edition 6; TNM 7, TNM classification, edition 7.

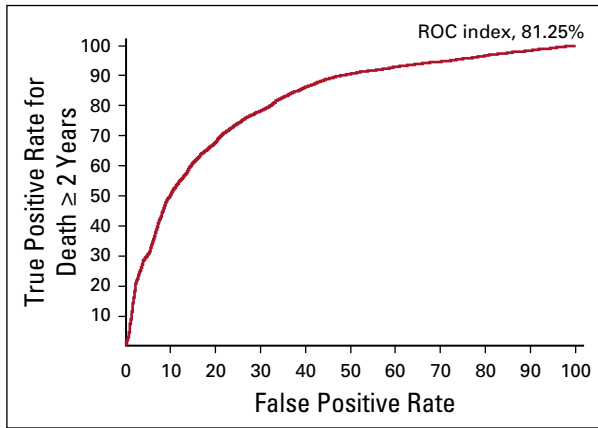


FIG 2. Area under the curve for 2-year survival (0.81), computed using BayesiaLab, and the test set comprising all editions. ROC, receiver operation characteristic.

≥ 2-year survival was 0.81, determined using BayesiaLab, and is shown in Figure 2. Model calibration was computed and is shown in Figure 3.

DISCUSSION

Mapping class labels requires that each label in the source edition (classification system) can be translated to a label in the target edition. Generally, this is not possible going from coarser to more granular classification systems (eg, from the 6th to the 7th edition of the TNM classification system).

In these situations, additional variables are required: either for use *in conjunction with* the original classification or to fully stage a tumor in the target edition. These additional variables are frequently unavailable or obtaining them comes with great cost.

We experienced these issues first hand. When creating the ground truth for evaluating predictive performance, only a relatively small subset of the total test set (12%; patients with a year of diagnosis between 2015 and 2016) could be considered for manual, rule-based reclassification. Even then, reclassification of the T descriptor was not possible in 23% of the patients in the subset. Specifically, we could not determine when a T3 in TNM7 would be a T2 in TNM6 because this required information about the presence of invasion into nearby anatomic structures. Similarly, we could not determine the 6th edition T and M descriptors for 7th edition nonmetastatic, multifocal T4 tumors because this required knowledge of the location(s) of the additional tumor nodules. Therefore, we did not evaluate performance of predicting individual descriptors.

On a more aggregated level, we could not fully determine the 6th edition stage group in 200 patients (5.6%). Although a relatively small number of records was involved, we thought this might still bias the test set. To investigate the potential effect of this bias, we performed an additional analysis. We looked at the (in real-life impossible) worst-case scenario by assuming that all 174 records where we could not decide between 2 stages (eg, stage IB or IIB)

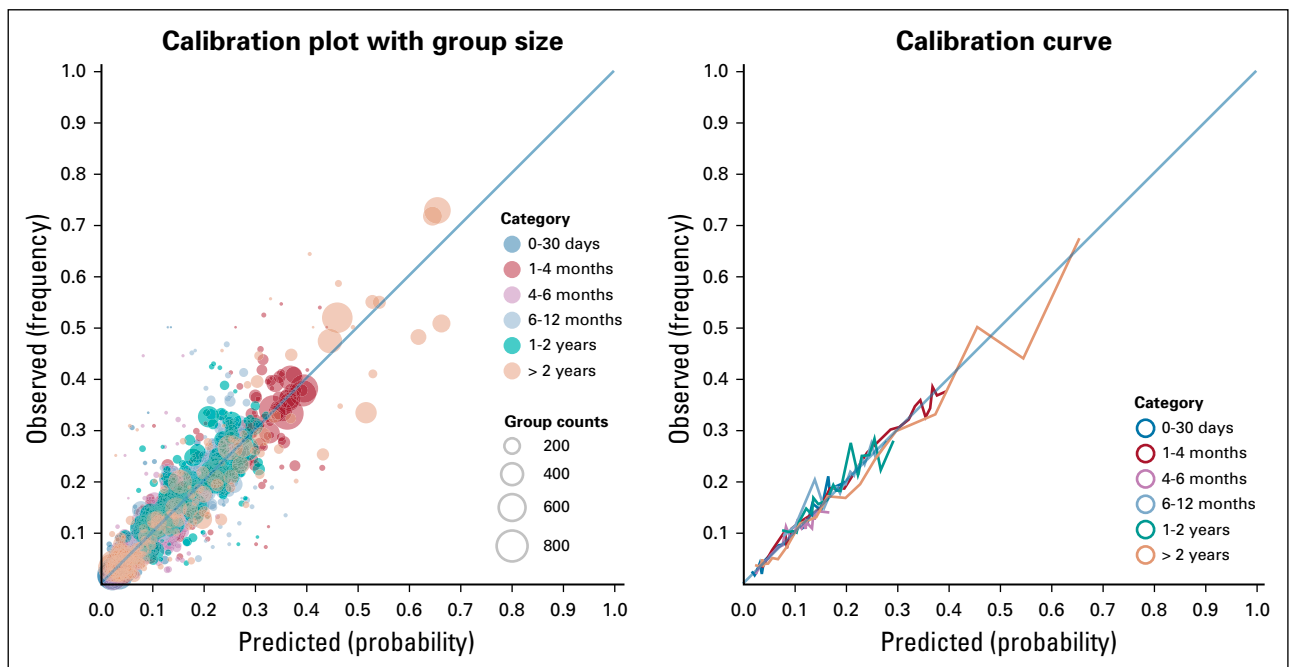


FIG 3. Visualization of model calibration for predicting survival. The calibration (bubble) plot shows, for each subpopulation, the relation between a survival category's predicted probability (x-axis) and its observed frequency in the dataset (y-axis). The size of each bubble corresponds to the population's size in the dataset. The calibration curve shows the same relation, but averages predicted/observed values by applying quantile-based binning with 1,000 samples per bin (21 bins).

would *always* be incorrectly classified, essentially doubling the error. After modifying the confusion matrices in this way, the averaged statistics were recalculated. Micro-averaged and macro-averaged specificity and accuracy were almost unaffected (values changed from 0.99 and 0.99 to 0.98 and 0.96, respectively, that is, a maximum decrease of 0.03). Micro-averaged and macro-averaged sensitivity decreased by a maximum of 0.1.

When using the ground truth set, the BN performed well when reclassifying the TNM stage group between editions: when predicting the 6th edition to the 7th edition data, all aggregated statistics (sensitivity, specificity, and accuracy) were ≥ 0.91 . Also, the fact that macro-averaged and micro-averaged statistics were close together implies the model is relatively insensitive to class imbalances. A similar observation can be made for predicting the 7th edition using 6th edition data. The results were only slightly worse, which is to be expected, considering the network has to predict a more granular output from a coarser input. Still, the model had an accuracy of 98%.

Performance for predicting 2-year survival can be considered more than adequate, especially considering the limited number of variables used for making the prediction and the number of possible outcome classes. Adding variables could help, but would come with additional complexity of the model. However, clinical decisions are not just based on the most likely outcome: probability is considered as well. Therefore, binary prediction (eg, 2-year survival: yes/no) seems inadequate to support decision making. Moreover, using a BN in such a way ignores one of its major strengths: the fact that it can communicate (conditional) probabilities. The calibration curves show that the model is well calibrated and that the estimated probabilities are close to the real probabilities.

Inspection of the confusion matrix reveals 4 situations where the model had difficulty in predicting the correct stage group, all related to possible stage shifting. When predicting the 6th edition stage group from 7th edition data, the errors stemmed from difficulty in handling patients where the input (node T_7) is T3. On inspection of the BN, it seemed the model predicted a T3 in the 7th edition to be a T2 with 58% probability and a T3 with 35% probability in the 6th edition. This explains the majority of mistakes made when the predicted descriptor should have been T3. The

original priors for this relation were set close to 50-50 (ie, a T3 becoming a T2 or T3 with equal probability) before applying EM learning; therefore, the change in probabilities appears to be an effect of optimizing the relationship between nodes T_567 and death. Additionally, the BN estimates the probability of a T3 in TNM7 being multifocal and thus a T4 in TNM6 to be fairly small, at 4.2%. Even if the actual percentage in the dataset is larger, the value T4 (in TNM6) would never be predicted, because the most probable outcome was selected as prediction.

In predicting the 7th edition stage group from 6th edition data, most of the misclassifications can be explained by the observation that a TNM6 T2 becomes either a T2a, T2b, or T3. Without additional information (ie, tumor diameter), it is not possible to be 100% accurate. Similarly, a T4 in TNM6 can become a T3 or T4, depending on multifocality of the tumor.

Even when using training data with classifications in both the 6th and 7th editions, the BN would not have been able to make any of these distinctions with certainty, because additional information is needed. However, like with predicting survival, the probabilistic reclassification we applied does not need to yield a discrete result: it is possible to assign a probability to each possible outcome, essentially creating a weighted reclassification. This is especially useful when reclassifying large datasets like the NCR.

We conclude that we have successfully created a BN that can aid in determining the TNM stage group of the 6th edition using 7th edition data, and vice versa, by using a training set that does not hold the known classifications in multiple editions but does hold survival to aid in the classification. Knowledge about changes between the editions of the classification system was successfully incorporated by modeling these changes as priors in the CPTs. The model parameters were estimated from data and therefore depend on specific distributions found in the Netherlands. However, considering NSCLC diagnostics, treatment, and survival are comparable in Western/developed countries, we expect the BN can be applied here as is, although validation would be required. This process is likely to work for other tumors and/or editions of the TNM classification system, but additional research is needed to establish generalizability.

AFFILIATIONS

¹Department of Research and Development, Netherlands Comprehensive Cancer Organization, Utrecht, the Netherlands

²Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, the Netherlands

CORRESPONDING AUTHOR

Melle S. Sieswerda, MD, MSc, Godebaldkwartier 419, 3511 DT Utrecht, the Netherlands; e-mail: m.sieswerda@iknl.nl.

AUTHOR CONTRIBUTIONS

Conception and design: All authors

Data analysis and interpretation: All authors

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Melle S. Sieswerda

Employment: United Neuroscience (I)

Inigo Bermejo

Employment: Cochlear

Dirk De Ruyscher

Consulting or Advisory Role: Bristol-Myers Squibb (Inst), AstraZeneca (Inst), Merck Serono (Inst), Celgene (Inst), Genentech (Inst), Seattle Genetics (Inst), MSD Oncology (Inst)

Research Funding: Bristol-Myers Squibb (Inst), AstraZeneca (Inst), Philips Research (Inst), Olink Bioscience (Inst)

André L. A. J. Dekker

Employment: Medical Data Works

Stock and Other Ownership Interests: Medical Data Works

Honoraria: Varian Medical Systems, Accuray, Roche, Janssen-Cilag, Philips Research (Inst), Varian Medical Systems (Inst)

Consulting or Advisory Role: Varian Medical Systems

Research Funding: Varian Medical Systems (Inst), Philips Healthcare (Inst), OncoRadiomics (Inst)

Patents, Royalties, Other Intellectual Property: Royalties from Mirada Medical on a deep-learning application in medical imaging (Inst), royalties from Varian Medical Systems on dosimetry software for radiation oncology (Inst), royalties from Health Innovation Ventures on radiomics, eLearning, and prediction models in cancer (Inst), royalties from PXI for software related to small animal irradiation for life sciences research (Inst), patents held currently: monitoring respiration based on plethysmographic heart rate signal. US patent 6,702,752.-Apparatus and method for monitoring respiration with a pulse oximeter. US patent 6,709,402, Monitoring mayer wave effects based on a photoplethysmographic signal. US patent 6,805,673, Monitoring physiologic parameters based on variations in a photoplethysmographic baseline signal. US patent 6,896,661, Monitoring physiologic parameters based on variations in a photoplethysmographic signal. US patent 7,001,337, Knowledge-based interpretable predictive model for survival analysis. US Patent 8,078,554, Dose distribution modeling by region from functional imaging. US Patent 8,812,240, Systems, methods and devices for analyzing quantitative information obtained from radiological images. US Patent 9721340 B2

Expert Testimony: Thoratec

Travel, Accommodations, Expenses: Medical Data Works, Varian Medical Systems, Roche, Janssen-Cilag

No other potential conflicts of interest were reported.

REFERENCES

- Greene FL, Sobin LH: The staging of cancer: A retrospective and prospective appraisal. *CA Cancer J Clin* 58:180-190, 2008
- Amin MB, Edge S, Greene F, et al (eds): *AJCC Cancer Staging Manual* (ed 8). New York, NY, Springer International Publishing, 2017
- Goldstraw P, Crowley J, Chansky K, et al: The IASLC Lung Cancer Staging Project: Proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol* 2:706-714, 2007
- Edge SB, Byrd DR, Compton CC, et al (eds): *AJCC Cancer Staging Manual* (ed 7). New York, NY, Springer, 2010
- Boffa DJ, Deterbeck FC, Smith EJ, et al: Should the 7th edition of the lung cancer stage classification system change treatment algorithms in non-small cell lung cancer? *J Thorac Oncol* 5:1779-1783, 2010
- Pearl J: *Causality: Models, Reasoning and Inference* (ed 2). New York, NY, Cambridge University Press, 2009
- Koller D, Friedman N: *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, MIT Press, 2009
- Conrady S, Jouffe L: *Bayesian Networks & BayesiaLab: A Practical Introduction for Researchers*. Franklin, TN, Bayesia USA, 2015
- Dempster AP, Laird NM, Rubin DB: Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 39:1-38, 1977
- Sokolova M, Lapalme G: A systematic analysis of performance measures for classification tasks. *Inf Process Manage* 45:427-437, 2009
- Perez F, Granger BE: IPython: A system for interactive scientific computing. *Comput Sci Eng* 9:21-29, 2007
- McKinney W: Data structures for statistical computing in Python. Presented at the 9th Python in Science Conference, Austin, TX, June 28-30, 2010. <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>
- Oliphant TE: *Guide to NumPy* (ed 2). Austin, TX, Continuum Press, 2015
- Pedregosa F, Varoquaux G, Gramfort A, et al: Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825-2830, 2011
- Niculescu-Mizil A, Caruana R: Predicting good probabilities with supervised learning. <https://www.cs.cornell.edu/~alex/papers/calibration.icml05.crc.rev3.pdf>

