Big data and interpretable models for outcome prediction in radiation oncology

BIG DATA AND INTERPRETABLE MODELS FOR OUTCOME PREDICTION IN RADIATION ONCOLOGY

Biche Osong

BIG DATA AND INTERPRETABLE MODELS FOR OUTCOME PREDICTION IN RADIATION ONCOLOGY

DISSERTATION

to obtain the degree of Doctor at the Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Pamela Habibović,
in accordance with the decision of the Board of Deans,
to be defended in public,
on Tuesday 7 March 2023 at 16.00 hrs.

by

Akuli Biche Akuli Biseh Osong

Supervisor:
Prof. dr. ir. Andre Dekker

Co-supervisors:
Dr. Johan van Soest
Dr. Inigo Bermejo Delgado

Assessment Committee
Prof. dr. Dirk De Ruysscher (Chair)
Dr. Timo Deist (ASML)
Prof. dr. Chris Terhaard (University Medical Center Utrecht)
Prof. dr. Anna Wilbik

TABLE OF CONTENTS

Part I
Mise en scene

CHAPTER 1

GENERAL INTRODUCTION AND THESIS OUTLINE

Introduction

Cancer refers to any one of many diseases characterized by the development of abnormal cells that divide uncontrollably. These cells continue to divide and grow and, at long last, overrun the healthy cells gaining the ability to infiltrate and destroy normal body tissue and sometimes spread to other parts of the body different from where they originated. These overgrown cells do not develop into healthy tissue and do not function normally. This uncontrollable growth gives rise to tumors that interfere with the human system's normal functioning, which can be lethal. Today, millions of people live with cancer or have had cancer at some point in their life, making the disease the second-leading cause of death worldwide for years (1–4), and since 2005, the preeminent cause of death in the Netherlands[1] (5). Figure 1.1 depicts the annual (2017) number of deaths by cause in the Netherlands (6), which shows cancer burden on the human population.

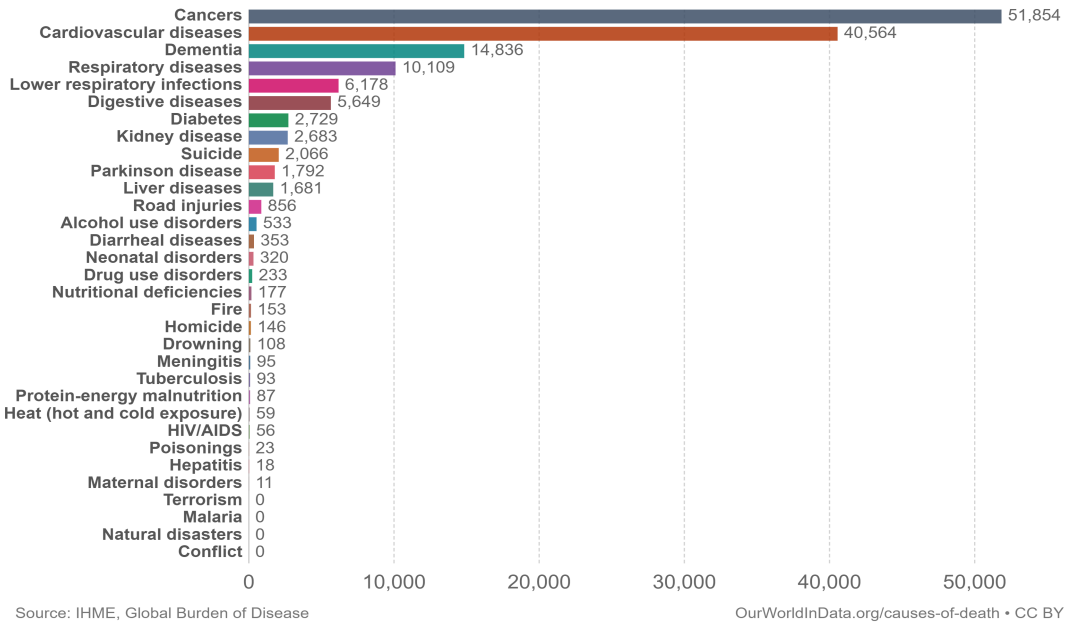

Figure 1.1. *The number of deaths by cause, the Netherlands, 2017.*

Over the past several decades, a whole field of research has opened up with a primary focus on understanding and finding the optimal treatment for this deadly disease challenging humans' health worldwide. In the 2002 – 2003 fiscal year, 1.43 billion euros was spent on cancer research in Europe (7) from public funding alone, and just ten years later, this figure stands at approximately 7.6 billion euros (8). These figures are even higher in the United States (7, 9).

The rest of this chapter is organized into five sections as follows. First, a brief introduction of the origin of radiation oncology and the science behind its therapeutic nature is discussed. The preceding two sections introduce Big Data in radiation oncology and machine learning

[1] Statistics Netherlands (CBS)

with clinical applicability. The penultimate section describes model explainability and interpretability to highlight the difference, while the last section gives the structure and blueprint of this thesis.

Radiation Oncology

The word oncology holds two meanings and has its roots in ancient Greek. The prefix "*onco-*" came from the Greek word "ogkos" which means lump, bulk, mass or tumor in modern Latin. The suffix "*-logy*" means study from the Greek word "logos". Therefore the term oncology means the study of tumors, i.e., the branch of science that deals with the study of tumors and cancers. Radiation, on the other hand, is the emission or transmission of energy in the form of particles via space or material medium, which can either be ionizing or non-ionizing depending on the characteristics of the particles.

Ionization is a process where electrons are being removed from atoms or molecules' orbitals by high-energy radiations such as Gamma rays, X-rays, and the higher ultraviolet part of the electromagnetic spectrum. This ionization causes chemical bonds between atoms or molecules to be broken, leaving them with unpaired electrons, called (free) radicals (10). These free radicals are very reactive and cause DNA damage which induces cellular senescence and stops cell division and proliferation processes. Generally, cells can recover from this kind of damage via cellular repair mechanisms, and their ability to recover from radiation damage largely depends on the cell type, with healthy cells recovering faster than cancerous cells (11). However, when the severity of the damage exceeds the cell's ability to repair itself, it ceases to carry out its biological functions and eventually dies (12).

This characteristic difference between normal and cancerous cells is being utilized in radiation oncology as a therapeutic option by administering a calculated amount of radiation dose to regions of the body where cancerous cells are targeted in multiple fractions over several days. The administered radiation dose is focused so that the cancerous cells receive most of the radiation dose, high enough to damage their DNA leading to cellular death or stopping them from further growing, but low enough for the healthy tissues far from the cancerous cell to avoid any complications such as radiation toxicity or sickness (13). Radiation is used as a therapy to get rid of all the cancerous cells ("curative treatment intent") and prevent them from recurrence or delaying their growth, and ease symptoms ("palliative treatment intent"). Radiation therapy is often given in combination with other treatments such as surgery, chemotherapy, hormone therapy, immunotherapy, or some mixture of these therapies depending on the tumor type, location, stage, and general health of the patient (14, 15).

Big Data

More than 50% of people diagnosed with cancer receive radiation therapy at some point in their treatment, either as a monotherapy or in combination with other therapies (13). Over the past several years, a continuous effort has been employed to improve and automate the cancer treatment process using information technology (IT) (16). The standard process of modern radiation treatment generates an abundance of data electronically stored across various disciplines such as radiation oncology, radiology, and other disciplines involved in the patient's care path. Patient information is mainly stored on platforms like the electronic medical record (EMR) system, picture archiving and communication system (PACS), and departmental information systems such as oncology information systems (OIS) or treatment planning systems (TPS). It usually comes in a variety of (un)structured formats such as the Digital Imaging and Communications in Medicine (DICOM) standard, relational databases, and free text. This sea of patient information has led to oncology's revolutionary Big Data era.

Machine learning

The digitalization of healthcare has significantly advanced the healthcare industry within the last decades. Innovations in this domain have led to a large volume of patient diagnosis, planning, and treatment data to be captured and stored in (un)structured electronic formats worldwide. Ideally, caregivers, stakeholders, and patients need these data to be translated into knowledge to assist them in their decision-making process (17, 18). The quest for evidence-based decision-making has encouraged the utilization of machine learning algorithms in healthcare (19). Machine learning, a sub-field of artificial intelligence (AI), studies the design of algorithms and their ability to learn from data and improve its performance through experience without being explicitly programmed (19, 20). Machine learning uses the theoretical knowledge from statistics to build mathematical models capable of detecting patterns within a dataset. In radiation oncology, machine learning can be employed on routine clinical data or images to elucidate us about a disease or system of interest. This insightful and actionable information provided by these models, makes them very valuable in cancer treatment and management.

Prediction models

Prediction models are mathematical formulas that can be used to forecast future outcomes. They are developed by processing large volumes of historical information in order for the model to learn how the different variables of data are related to each other and their association with the outcome of interest. The process of developing these models involves four main steps. First, the data is collected, explored using exploratory data analysis techniques, and cleaned for data inconsistency, outliers, and missing information (21, 22).

Secondly, the model is trained on part of the collected data using statistical methods. Thirdly, the model is tested on the other part of the data not used to train the model, though in some cases, like in cross-validation, the training and testing datasets are changed to determine model robustness. Lastly, the model is validated on an external dataset to estimate the model's performance in a new setting and, as such, an essential step before using a model in clinical practice (23). In radiation oncology, prediction models are used to calculate the probability of a given patient experiencing a particular outcome or event (survival, disease recurrence, adverse treatment effects...) of interest depending on the patient-specific values of the model's predictive variables. They are also used to select patients who might not benefit from a specific treatment like proton therapy (24–26). Such insights from these prediction models could guide the selection of the most appropriate treatment for a patient or modify their treatment options to improve outcomes (27).

Regression models

Over the last decade, regression methods have become an indispensable component of predictive modeling tasks in most fields. They are concerned with describing the relationship between a response variable (also called output, outcome, or dependent variables) and one or more explanatory variables (also called independent, predictor, or input variables). For example, equation 1.1 is a typical representation of a regression model where the variable(s) on the right-hand side of the equation is said to have an (causal) effect on the outcome variable on the left-hand side of the equation (28–30).

$$Outcome = \beta_0 + \beta_1 * Variable_1 + \beta_2 * Variable_2 + \cdots + \beta_n * Variable_n + \epsilon \tag{1.1}$$

The β values are multiplicative factors called regression coefficients, and they indicate the magnitude of influence the normalized predictor or variable has on the outcome (31). These β values can serve as a means to create risk groups for the outcome variable of interest, as implemented in chapters 4 and 5 (32, 33). The outcome variable is often continuous, discrete, or a count, which might slightly modify the representation of equation 1.1. Generally, there are several regression models, but this thesis is limited to the regression method applied to time-to-event data. The time-to-event data format is essential, especially in healthcare, since caregivers may wish to know the probability of a patient developing an event of interest and when that said event will occur. Therefore, unlike popular regression methods like linear and logistic, the outcome of interest in time-to-event analysis has two unique parameters: the time and event factor. The Cox proportional hazard regression model is the go-to model for analyzing time-to-event data in most fields, especially medical research. However, other non-parametric methods like decision trees are being used lately see chapter 7.

Decision Trees

A decision tree is a tree-like (roots up) graphical structure that depicts a set of decisions and every potential outcome or result of making those decisions. Decision trees are non-linear methods that iteratively use the independent variable, which has the strongest association with the dependent variable, to split a population into progressively smaller subgroups according to some specific criterion (34). The root node is the most important variable on the tree, and is positioned at the top of the tree, from where the rest of the nodes on the tree are linked. The intermediate nodes are all the subsequent variable(s) splits, and the leaf node is the tree's final node, which holds the prediction or insights needed. The tree branches link the nodes and provide the different options or courses of action available when making a particular decision.

Decision trees can be developed in various situations, from something simple and personal such as "should I go to the beach today?" to a more significant and complex scientific undertaking. An example is the decision tree deployed during the coronavirus disease 2019 (COVID -19) pandemic to support parents and childcare workers determine when a child is allowed/fit to attend school as long as corona measures apply (Figure S1.3 supplemental material[2]). Decision trees are less challenging to explain to end-users since they take a problem with multiple outcomes and display the solutions in a graphical, easy-to-understand, and straightforward format that shows the relationship between different decisions and their outcomes. By displaying a sequence of steps, this structured model enables the end-users to visualize and understand how and why each potential decision may lead to the next, using the "IF-THEN" mutually exclusive options given on the branches. In other words, end-users observe every conceivable outcome on the tree and weigh each course of action against the decision's risks and reward.

However, one crucial dilemma when building a decision tree is deciding when to stop growing the tree (no further splitting of the node) and use a specific tree as the final model. When the tree size is too big, there is a high probability that the tree will overfit the data and thus fail to generalize on an external population. If too small, some essential branches and intermediate nodes will be absent, leading to poor performance of the tree. Generally, most researchers either build a huge tree and select an appropriate subtree by pruning off branches or use a stopping rule during training to circumvent this dilemma (35, 36). In recent years, much attention has been given to an ensemble method called random forest, which combines several base decision trees to produce an optimal model with better predictive performance. However, this method defeats one of the desired characteristics of decision trees, which is to provide insights into the data by graphically displaying the results in a manner readily understood and interpreted by end-users.

[2] https://www.boink.info/beslisboom

Bayesian Network

Bayesian Networks (BN) are probabilistic graphical models (PGM) or structured probabilistic models that can represent and reason in uncertain domains such as healthcare. This hybrid model from probability and graph theory visually describes the relationship between a set of random variables and their conditional dependencies. These relationships are represented in a directed acyclic graph (DAG), where each node in the network represents a variable (discrete or continuous), and directed edges, arcs, links, connections, or arrows denote causal relationships or dependencies between these variables (37). For example, an edge from variable A to B may represent a causal effect relationship of variable A on B.

Bayesian networks are very appealing in medicine because of their ability to make predictions with incomplete information and inference on any variable in the structure. Their graphical nature also enables the incorporation of prior knowledge and causal intervention, making them suitable decision-making tools with clear and understandable reasoning in their decision process (38, 39). Generally, the directed edges within Bayesian networks structures are generated by an algorithm using a statistical test to ascertain that variable A is probabilistically independent of variable B conditional on its parents (40, 41). However, this method does not always produce valid structures, especially in healthcare, since it might include relationships between variables where the direction of influence between variables is impossible. For example, having a variable being influenced by a variable that is collected later in the future (e.g., recurrence after two years causing neo-adjuvant chemotherapy yes/no) or having two unrelated variables influencing each other (e.g., age and gender).

Another option is to request a domain expert to specify the Bayesian network structure, which is also not optimal since the expert might be biased based on their knowledge and experience of the domain. One way to improve this expert-specified structure limitation is by surveying opinions from multiple experts with the assumption that arcs specified by most experts will likely be the ground truth or accurate representation of the domain (majority voting theory as implemented in chapter 8).

Explainability and Interpretability

The use of machine learning models as decision support tools has been steadily rising in different sectors. With the arrival of the General Data Protection Regulation (GDPR), which grants end-users the "right to explanation" of algorithmic decisions (42–44), an indirect burden is placed on prediction models emphasizing accountability and transparency. End-users can, therefore, be interested in both the predicted outcome and the interpretability of the prediction model. Trusting the system powered by these machine learning models

will require a better understanding of how the system arrived at a particular prediction. However, the extent of one's understanding of these systems largely depends on whether the model is interpretable or explainable. Though explainability and interpretability are different, in the context of machine learning, they are separated by a thin line, with most authors using the terms interchangeably (45, 46). While they might be closely related, they are defined differently.

> *"The interpretability of a model relates to how a human can observe the cause and effect within a system. To rephrase, it is the degree to which one can consistently understand the cause of the model's decision. That is seeing what happens when the parameters of the model change" (46–49).*

> *"The explainability of a model, on the other hand, is the extent to which the model's internal mechanics can be explained in human terms. That is if the developers of the model can explain why the model arrived at a specific decision and can pass the interpretation down to users" (45, 46, 50)*

The subtle difference between interpretability and explainability could easily be missed. In two words, interpretability is model discernment, and explainability is model explication. Let's take a simple chemistry experiment on titration to illuminate this difference further. The steps taken until the color change in the solution is observed can be considered interpretability. However, the chemistry behind the experiment (reaction between the reactants) is the definition of explainability. Model explainability is not discussed further beyond this point because this thesis's primary focus is model interpretability. Figure 1.2 shows some commonly used machine learning models in radiation oncology (51) in order of increased interpretability.

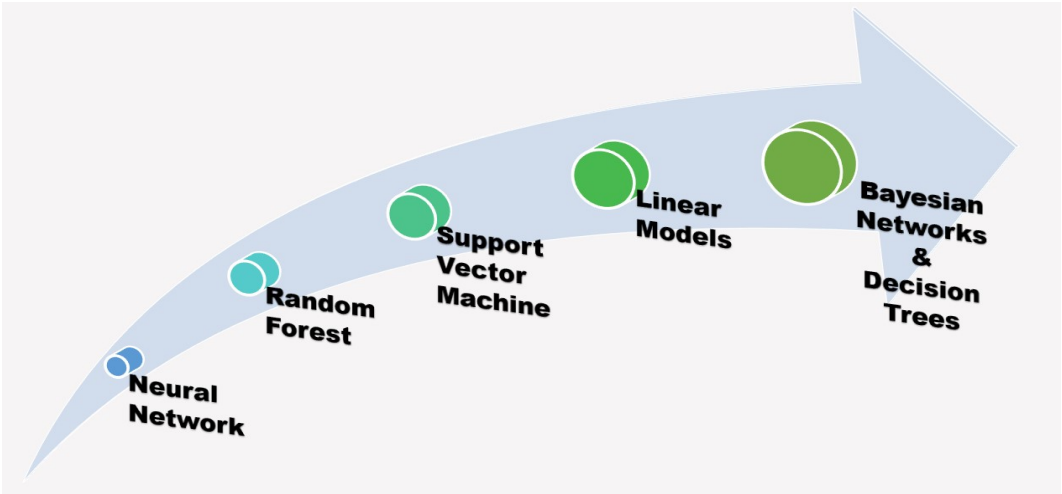

Figure 1.2. *Commonly used machine learning models in radiation oncology in order of increased interpretability.*

Regardless of whether or not there is any regulation or restriction on prediction models in terms of accountability and transparency, it is pivotal for prediction models to have a certain level of interpretability to build trust with end-users and ease communication of their workings to stakeholders. By definition, interpretable models should be inherently easier to translate to a less technical audience than non-interpretable models.

Regression is one technique that produces very interpretable models, though they might sometimes lose their interpretability, especially when they contain splines or interplay of explanatory variables. Informative graphical tools which translate complex models to a visual form are often used in such situations to assist end-users and stakeholders in understanding the rationale and results of the model's predictions. Nomograms are often the go-to translational tool used to accommodate models with increased complexity and provide visuals to facilitate comprehension. Other methods like decision trees, which do not need translation since they are graphical structures with an IF-THEN flowchart form, are also interpretable. Though they can become cumbersome to interpret when grown too big, pruning off some branches helps keep them interpretable and straightforward.

In healthcare and especially radiation oncology, where machine learning models are significant assets for better cancer care, deployed models need to have some level of interpretability. Deployed models that lack interpretability face plenty of challenges with their potential and impact greatly limited due to lack of trust and fairness (52, 53), given that accountability and transparency are pivotal for end-users. Interpretable models like Bayesian networks can elevate trust and fairness amongst end-users since they can be elicited by domain experts (54, 55), who are also the end-user in most cases. Their graphical nature, which captures the conditional dependency between the variables, makes them transparent.

Irrespective of model choice, understanding how a machine learning model works and including end-users throughout the model-building process helps align the researcher's activities with the end-users' vital questions and the organization's needs. Therefore, this thesis aims to contribute to scientific knowledge of Big Data and interpretable machine learning models for outcome prediction in radiation oncology.

Structure of the thesis

This thesis contains four main sections, as described in table 1.1. Section 1 provided a general introduction to this thesis. Section 2 provides some basic theory on Big Data concepts, while the analyses section contains original research on prediction models in radiation oncology. The last section concludes this thesis with future perspectives and recommendations for building interpretable machine learning models in healthcare.

Table 1.1. *Summary of the topics and characteristics of the studies presented in the different chapters.*

Section	Part	Chapter	Study Type	Disease	Outcome
Introduction		1	-	-	-
Theory	Big Data	2	-	-	-
		3			
Analysis	Regression	4	Development & internal validation	Cervical Cancer	Progression-free and overall survival
		5	Development & internal validation	Spinal Metastases	Overall Survival
	Decision Tree	6	Development & internal validation	Multiple Cancer Types	Radiotherapy Compliance
		7	Development & external validation	Cervical Cancer	Overall Survival
	Bayesian network	8	Development & internal validation	Rectal Cancer	Tumor Recurrence
		9	Development & internal validation	Lung Cancer	Two Years Survival
Discussion		10	-	-	-

The part dedicated to Big Data, a literature review section, comprises two chapters. Chapter 2 paints a picture of Big Data in healthcare and its benefits for better cancer care. Chapter 3 describes the use of Big Data in radiation oncology and provides some of the challenges involved in bringing Big Data technology into practice and some solutions.

Section 3 is split into three parts based on the analytical technique, with each part having two chapters. In the regression part, chapter 4 describes the development and validation of a multivariable Cox proportional hazard regression model to predict overall and progression-free survival for cervical cancer patients receiving chemoradiation. Chapter 5 translates a multivariable Cox proportional hazard regression model with an interaction term to a nomogram. The nomogram can predict 1, 2, and 6 months' overall survival for patients with metastatic spinal tumors, and its visual form could serve as a decision support tool for the personalized management of these patients.

Chapter 6 in the decision tree part describes the development of a decision tree to predict radiotherapy compliance for elderly cancer patients. Chapter 7 extends the decision tree

methodology to survival data to predict overall survival for stage IIB-IVA cervical cancer patients with squamous cell carcinomas.

The last part of section 3 concerns Bayesian network models. In chapter 8, an expert elicited Bayesian network to predict tumor recurrence in rectal cancer patients was compared with an algorithmic Bayesian network on a cohort of 6754 patients from 14 different international clinical trials in Europe. In chapter 9, we used knowledge from chapter 8 (expert structures are more interpretable but less predictive while algorithmic networks are more predictive but less interpretable) to develop a hybrid Bayesian network that involves the interplay between expert knowledge and a hill-climbing algorithm to predict two-year survival in non-small cell lung cancer patients.

Section 4 contains a single chapter (Chapter 10) that discusses the importance and challenges of interpretable machine learning prediction models in radiation oncology and concludes this thesis.

Supplemental materials

Figure 1.3. *The decision tree deployed during the coronavirus disease 2019 (COVID -19) pandemic to support parents and childcare workers determine when a child is allowed to attend school as long as corona measures apply in the Netherlands.*

Bibliography

[1] Robert T Greenlee, Taylor Murray, Sherry Bolden, and Phyllis A Wingo. Cancer statistics, 2000. *CA: a cancer journal for clinicians*, 50(1):7–33, 2000.

[2] Paola Pisani, D Maxwell Parkin, Freddie Bray, and Jacques Ferlay. Estimates of the worldwide mortality from 25 cancers in 1990. *International journal of cancer*, 83(1):18–29, 1999.

[3] Ahmedin Jemal, Freddie Bray, Melissa M Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: a cancer journal for clinicians*, 61(2):69–90, 2011.

[4] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.

[5] OECD, European Observatory on Health Systems, and Policies. *Netherlands: Country Health Profile 2017*. 2017. doi: https://doi.org/https://doi.org/10.1787/9789264283503-en. URL https://www.oecd-ilibrary.org/content/publication/9789264283503-en.

[6] Hannah Ritchie and Max Roser. Causes of death. *Our World in Data*, 2018. https://ourworldindata.org/causes-of-death.

[7] Seth Eckhouse and Richard Sullivan. A survey of public funding of cancer research in the european union. *PLoS Med*, 3(7):e267, 2006.

[8] Mursheda Begum, Grant Lewison, Mark Lawler, and Richard Sullivan. Mapping the european cancer research landscape: An evidence base for national and pan-european research and funding. *European Journal of Cancer*, 100:75–84, 2018.

[9] Janet Fricker. Cancer research funding in europe. *Molecular oncology*, 1(2):131–134, 2007.

[10] Michael Douglass. Eric j. hall and amato j. giaccia: Radiobiology for the radiologist, 2018.

[11] Ankush Gosain and Luisa A DiPietro. Aging and wound healing. *World journal of surgery*, 28(3):321–326, 2004.

[12] Margaret M Mc Gee. Targeting the mitotic catastrophe signaling pathway in cancer. *Mediators of inflammation*, 2015.

[13] Susan Hall, Santosh Rudrawar, Matthew Zunk, Nijole Bernaitis, Devinder Arora, Catherine M McDermott, and Shailendra Anoopkumar-Dukie. Protection against radiotherapy-induced toxicity. *Antioxidants*, 5(3):22, 2016.

[14] AC Society. Cancer treatment & survivorship facts & figures 2019–2021, 2019.

[15] Kimberly D Miller, Leticia Nogueira, Angela B Mariotto, Julia H Rowland, K Robin Yabroff, Catherine M Alfano, Ahmedin Jemal, Joan L Kramer, and Rebecca L Siegel. Cancer treatment and survivorship statistics, 2019. *CA: a cancer journal for clinicians*, 69 (5):363–385, 2019.

[16] Steven B Clauser, Edward H Wagner, Erin J Aiello Bowles, Leah Tuzzio, and Sarah M Greene. Improving modern cancer care through information technology. *American journal of preventive medicine*, 40(5):S198–S207, 2011.

[17] Cary Oberije, Georgi Nalbantov, Andre Dekker, Liesbeth Boersma, Jacques Borger, Bart Reymen, Angela van Baardwijk, Rinus Wanders, Dirk De Ruysscher, Ewout Steyerberg, et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. *Radiotherapy and Oncology*, 112(1):37–43, 2014.

[18] Issam El Naqa, Dan Ruan, Gilmer Valdes, Andre Dekker, Todd McNutt, Yaorong Ge, Q Jackie Wu, Jung Hun Oh, Maria Thor, Wade Smith, et al. Machine learning and modeling: Data, validation, communication challenges. *Medical physics*, 45(10): e834–e840, 2018.

[19] Alexander FI Osman. Radiation oncology in the era of big data and machine learning for precision medicine. In *Artificial Intelligence-Applications in Medicine and Biology*. IntechOpen, 2019.

[20] Alberto Traverso, Frank JWM Dankers, Biche Osong, Leonard Wee, and Sander MJ van Kuijk. Diving deeper into models. *Fundamentals of clinical data science*, pages 121–133, 2019.

[21] John W Tukey et al. *Exploratory data analysis*, volume 2. Reading, Mass., 1977.

[22] Frederick Hartwig and Brian E Dearing. *Exploratory data analysis*. Number 16. Sage, 1979.

[23] Gary S Collins, Joris A de Groot, Susan Dutton, Omar Omar, Milensu Shanyinde, Abdelouahid Tajar, Merryn Voysey, Rose Wharton, Ly-Mee Yu, Karel G Moons, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC medical research methodology*, 14(1):1–11, 2014.

[24] Alexander R Delaney, Max Dahele, Jim P Tol, Ingrid T Kuijper, Ben J Slotman, and Wilko FAR Verbakel. Using a knowledge-based planning solution to select patients for proton therapy. *Radiotherapy and Oncology*, 124(2):263–270, 2017.

[25] Jasper Kouwenberg, Joan Penninkhof, Steven Habraken, Jaap Zindler, Mischa Hoogeman, and Ben Heijmen. Model based patient pre-selection for intensity-modulated proton therapy (impt) using automated treatment planning and machine learning. *Radiotherapy and Oncology*, 158:224–229, 2021.

[26] Makbule Tambas, Roel JHM Steenbakkers, Hans P van der Laan, Atje M Wolters, Roel GJ Kierkels, Dan Scandurra, Erik W Korevaar, Edwin Oldehinkel, Tineke WH van Zon-Meijer, Stefan Both, et al. First experience with model-based selection of head and neck cancer patients for proton therapy. *Radiotherapy and Oncology*, 151:206–213, 2020.

[27] Lijing Lin, Matthew Sperrin, David A Jenkins, Glen P Martin, and Niels Peek. A scoping review of causal methods enabling predictions under hypothetical interventions. *arXiv preprint arXiv:2011.09815*, 2020.

[28] Alan O Sykes. An introduction to regression analysis. 1993.

[29] Michael A Golberg and Hokwon A Cho. *Introduction to regression analysis*. WIT press, 2004.

[30] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[31] Ewout W Steyerberg et al. Clinical prediction models, 2019.

[32] Biche Osong, Ilknur Sanli, Paul C Willems, Leonard Wee, Andre Dekker, Seok Ho Lee, and Johan van Soest. Overall survival nomogram for patients with spinal bone metastases (sbm). *Clinical and translational radiation oncology*, 28:48–53, 2021.

[33] Youn Ji Kim, Young Saing Kim, Jin Woo Shin, Biche Osong, and Seok Ho Lee. Prediction scoring system based on clinicohematologic parameters for cervical cancer patients undergoing chemoradiation. *International Journal of Gynecologic Cancer*, 30(11), 2020.

[34] Gordon S Linoff and Michael JA Berry. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2011.

[35] Charles Elkan. The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*, volume 17, pages 973–978. Lawrence Erlbaum Associates Ltd, 2001.

[36] James Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.

[37] Judea Pearl. Bayesian networks. 2011.

[38] Peter Lucas. *Bayesian networks in medicine: a model-based approach to medical decision making*. na, 2001.

[39] Emanuela Barbini, Pietro Manzi, and Paolo Barbini. Bayesian approach in medicine and health management. *Current Top Public Health*, 2013.

[40] Marco Scutari. Learning bayesian networks with the bnlearn r package. *arXiv preprint arXiv:0908.3817*, 2009.

[41] Marco Scutari and Jean-Baptiste Denis. *Bayesian networks: with examples in R*. CRC press, 2021.

[42] General Data Protection Regulation. Regulation eu 2016/679 of the european parliament and of the council of 27 april 2016. *Official Journal of the European Union. Available at: http://ec. europa. eu/justice/data-protection/reform/files/regulation_oj_en. pdf (accessed 20 September 2019)*, 2016.

[43] Ian Sample. Ai watchdog needed to regulate automated decision-making, say experts. *The Guardian*, 27:01–17, 2017.

[44] Andrew Burt. Is there a'right to explanation'for machine learning in the gdpr. *International Association of Privacy Professionals*, 1, 2017.

[45] Ričards Marcinkevičs and Julia E Vogt. Interpretability and explainability: A machine learning zoo mini-tour. *arXiv preprint arXiv:2012.01805*, 2020.

[46] Aniek F Markus, Jan A Kors, and Peter R Rijnbeek. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics*, 113:103655, 2021.

[47] Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38, 2019.

[48] Been Kim, Oluwasanmi Koyejo, Rajiv Khanna, et al. Examples are not enough, learn to criticize! criticism for interpretability. In *NIPS*, pages 2280–2288, 2016.

[49] Owen Shen. Interpretability in ml: A broad overview. *The Gradient*, 2020.

[50] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[51] Timo M Deist, Frank JWM Dankers, Gilmer Valdes, Robin Wijsman, I-Chow Hsu, Cary Oberije, Tim Lustberg, Johan van Soest, Frank Hoebers, Arthur Jochems, et al. Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers. *Medical physics*, 45(7):3449–3459, 2018.

[52] Zachary C Lipton. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57, 2018.

[53] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

[54] ME Kragt. A beginners guide to bayesian network modelling for integrated catchment management. landscape logic technical report no. 9. 2009.

[55] Jose M Pena. Finding consensus bayesian network structures. *Journal of Artificial Intelligence Research*, 42:661–687, 2011.

Part II
Big Data

CHAPTER 2

BIG DATA FOR BETTER CANCER CARE

***Authors**:*
***Biche Osong**, Andre Dekker, and Johan van Soest.*

***Adapted from**: British Journal of Hospital Medicine Vol. 80, No. 6*
DOI**: **10.12968/hmed.2019.80.6.304

Abstract

This editorial seeks to paint a clearer picture of Big Data in healthcare and the benefits for better cancer care. It gives a simple and widely used definition of Big Data and explains some of its pertinent characteristics (volume, velocity, variety, veracity) with respect to healthcare. A brief discussion of some solutions provided by Big Data for healthcare challenges together with a vivid examples where Big Data have been implemented to improve operational efficiency and clinical excellence for hospitals. Big Data has been gradually changing the face of healthcare, with the promise to improve cancer care, decision making in real time via decision support systems(DSS) and better operations in healthcare institutions. However, making Big Data Findable Accessible Interoperable and Reusable (FAIR) in healthcare would be an essential improvement not only to circumvent the data sharing problem, but also open up the possibility to use and reuse healthcare Big Data and trust the conclusions drawn from Big Data.

Introduction

Cancer is defined by the abnormal and uncontrollable growth of cells and is the second leading cause of death worldwide (1). The heterogeneity in disease manifestations, various treatment options, patients' preferences and deciding which treatment is optimal for an individual patient are some of the most significant challenges in health care.

To make reliable decisions, we need large amounts of data. According to IBM, approximately 2.5 quintillion bytes of digital data are generated from humans' daily activities (e.g. from groceries, demographic and administrative medical records) (2). The acquisition or extraction of these extensive and voluminous data is colloquially known as Big Data. There is no standard definition for Big Data, but in simplistic terms, it is a significantly large and complex dataset, which is impossible to adequately manage and process with traditional software (3). However, Gartner defines Big Data as:

> "*Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation*".

Although Big Data is perceived by many as just an extensive collection of datasets, there is more to the definition of "big" than only the volume aspect.

Big Data Anatomy in Healthcare

Like the human body, Big Data has several distinct features. To have a better grasp of Big Data in healthcare, we can look at four aspects regularly used to describe Big Data.

Volume
With the advancements in medical technology, a considerable amount of patients' data can be generated from clinical practice (e.g. diagnostic and therapeutic procedure information, omics) and in public health (e.g. wearables and mobile devices).

Velocity
The switch from recording patients medical information on paper to an electronic health records systems (EHRs) have increased the amount of data which can be accessed and shared via a secure information system at any time. Hence, reducing the time between data generation and processing for real-time (or near real-time) decision making.

Variety
Healthcare providers use different media formats (numbers, images, videos, text or audio) which can be classified into three main groups (structured, semi-structured or unstructured)

Veracity

The quality and source of data are very pivotal to make the right inference. The mere fact that data comes from different sources opens the possibilities of bias, uncertainty, inconsistency, incompleteness or imprecision. All these affect the accuracy of the results and conclusions drawn from these heterogeneous sources.

Big Data in Healthcare

In 2006, Clive Humby, a UK mathematician, stated:

> "*Data is the new oil. It is valuable, but if unrefined it cannot be used.*"

This analogy is not without merit as data fuels most if not all industries (e.g. sports, agriculture, education). This quote and the accompanying hype is also influencing the health-care sector as it continuously strives to add value to patient care. These efforts can be split into two aims: to improve operational efficiency and to increase clinical insights and excellence (in terms of personalized or precision medicine).

Improve operational efficiency

Owing to the increased incidence of cancer worldwide, combined with rising health-care costs, operational efficiency in clinical practice becomes important. Optimizing or automating clinical workflows could reduce human workload, increasing the number of patients which can be treated and reducing the time spent receiving medical care.

Radiotherapy, one of the most widely used and effective cancer treatment options, is already using Big Data to automate image contouring and treatment planning tasks (4). This automation results in less variability in treatments, because it limits unnecessary human variation. In the end, Big Data and proper application of methods can result in a shift of repetitive and time-intensive tasks usually performed by humans to computer supervision.

Staff allocation per shift has been one of the challenges in clinical management. Finding the perfect balance between over-staffing departments (while providing satisfactory patient care) and under-staffing (with unsatisfactory care) is a challenging task. Big Data can help to solve this dilemma by estimating the expected number of patients visiting an emergency department on a daily and hourly base (5).

Clinical excellence

As our understanding of the human body increases, selecting the best treatment option for a particular patient becomes more complicated. The information at hand is much larger than the human brain can comprehend (6), resulting in treatment outcomes which are unpredictable by humans (7). Fortunately, Big Data can be used to predict treatment

outcomes, allowing humans to make decisions between various treatment options (and their outcomes). These predictions can be used to target patients who need rapid interventions or those will benefit from (neo)adjuvant treatments.

Eventually, applying such Big Data strategies could have benefits in terms of costs, quality and time. For example, knowing which patients do not need (neo)adjuvant treatment strategies would save costs for this treatment, improve quality of life (in terms of treatment burden or adverse treatment effects), and reduce treatment time.

This also results in the concept of value-based care (8), where the interventions are weighed against their actual benefit to the patient. As an example, in personalized treatment where the patient's specific tumor information is used to ascertain treatment response, Big Data can identify tumor-specific patterns information contained within the pixels of medical images which could assist the physician and radiologist in their diagnostics and decision-making process (9).

Impediments of Big Data in Healthcare

William Cowper said in the late 18^{th} century, "*Variety is the spice of life*". However, too much variety without a proper definition of terms can be detrimental for Big Data in health care. Big Data within this domain are coming from a variety of sources which when combined can give a richer insight or help provide better care. But the major challenge in data sharing among different sources within health care (both within and without health-care organizations) is how to preserve participants' privacy while still benefiting future patients. Data are currently stored in systems which are optimized for a specific (clinical) purpose, which generally does not target re-use by users who are not the data holder or vendor of the system (10).

To address these problems of data sharing in health care will require proper data descriptions with formal and well-defined ontologies (terminologies, properties and their respective relationships) within the domain. These ontologies are standardized in a machine-readable format (with human-readable representations) and shared among the different participants (10). The emphasis on having a common controlled vocabulary or standardizing an ontology in health care is essential to avoid participants being unable to use each other's data.

Benefits of FAIR Big Data in healthcare

To fully benefit from Big Data in health care, data and databases should firmly adhere to the FAIR principles (11), i.e. they should be findable, accessible, interoperable, reusable (FAIR) but still respecting patients' privacy and medical confidentiality. These principles do

not state that all data should be publicly available but urge that at least their descriptions (e.g. what kind of data are available) are published, including the means to contact the data (owners or maintainers), and which (semantic) representations are used. Using rich metadata descriptions of the actual data should improve data reuse and make data FAIR.

Conclusion

Big Data is an emerging field which has come to revolutionize the way we think and act in the health-care sector. Without a standard definition of the term, many have taken this to mean collecting a vast amount of data without looking at the true meanings of Big Data. The complexity of Big Data is an essential factor in health care, as data are coming from different sources, with accompanying different data privacy regulations, ownership and security implications.

Bibliography

[1] Wenpeng You and Maciej Henneberg. Cancer incidence increasing globally: the role of relaxed natural selection. *Evolutionary applications*, 11(2):140–152, 2018.

[2] Shivam Arora. Data science vs. big data vs. data analytics. *Elérhető: www. simplilearn. com/data-science-vs-big-data-vs-data-analytics-article (A letöltés dátuma: 2019. 03. 29.)*, 2019.

[3] J Deng. Big data in radiation oncology: challenges and opportunities. *Cancer Sci Res Open Access*, 1(2):1–2, 2014.

[4] Tim Lustberg, Johan van Soest, Mark Gooding, Devis Peressutti, Paul Aljabar, Judith van der Stoep, Wouter van Elmpt, and Andre Dekker. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiotherapy and Oncology*, 126(2):312–317, 2018.

[5] Bernard J Morzuch and P Geoffrey Allen. Forecasting hospital emergency department arrivals. In *26th Annual Symposium on Forecasting, Santander, Spain*, 2006.

[6] Amy P Abernethy, Lynn M Etheredge, Patricia A Ganz, Paul Wallace, Robert R German, Chalapathy Neti, Peter B Bach, and Sharon B Murphy. Rapid-learning system for cancer care. *Journal of Clinical Oncology*, 28(27):4268, 2010.

[7] Cary Oberije, Georgi Nalbantov, Andre Dekker, Liesbeth Boersma, Jacques Borger, Bart Reymen, Angela van Baardwijk, Rinus Wanders, Dirk De Ruysscher, Ewout Steyerberg, et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. *Radiotherapy and Oncology*, 112(1):37–43, 2014.

[8] Sarah Gentry and Padmanabhan Badrinath. Defining health in the era of value-based care: lessons from england of relevance to other health systems. *Cureus*, 9(3), 2017.

[9] RJ GGillies, PE Kinahan, and H Hricak. Radiomics: Images are more than pictures. *They Are Data. Radiology*, 278(2):563–577, 2016.

[10] Tim Lustberg, Johan van Soest, Arthur Jochems, Timo Deist, Yvonka van Wijk, Sean Walsh, Philippe Lambin, and Andre Dekker. Big data in radiation therapy: challenges and opportunities. *The British journal of radiology*, 90(1069):20160689, 2016.

[11] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. Andra waagmeester. *Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons*, 2016.

CHAPTER 3

APPLICATIONS OF BIG DATA IN RADIATION ONCOLOGY

***Authors**:*
***Biche Osong**, Andre Dekker, and Johan van Soest.*

***Adapted from**: The Modern Technology of Radiation Oncology, Volume 4, Chapter 11*
ISBN**: **978-1-951134-02-0

Abstract

Radiation therapy is a very complex process that consists of multiple steps. Thanks to technological advances, an enormous amount of data is generated from each patient during treatment and beyond. Collecting data from the treatment process is comparable to a snowball rolling down a hill. This rapid growth in the amount of electronically stored data from routine clinical practice has led to an explosion of interest in using Big Data techniques to address clinical challenges. With robust information technology infrastructure and decades of archived digital data, radiation oncology is well-positioned to take advantage of these Big Data techniques for better cancer care. Radiation oncologists also need help translating the available data generated across the continuum of care into knowledge to support decision-making in their clinical practice. This chapter gives a gentle introduction to Big Data and describes areas from which Big Data in radiation oncology stems. In addition, an overview of Big Data application in radiation oncology, limitations of application, and future perspective.

Introduction

Radiation oncology is a sub-specialty of oncology that uses ionizing radiations to treat abnormal and uncontrollable growing cells (neoplasia or lesions) with a curative (malignant cells localized to one area of the body) or palliative intent. This radiation treatment technique is called radiotherapy or radiation therapy and is often abbreviated RT, RTx, or XRT. Radiation therapy is a local treatment. It only focused on the region of the body where the malignant cells are located to deliver a precise amount of calculated dose of radiation to this region of the body with cancerous cells using a linear accelerator. The purpose is to eradicate these cells and improve quality of life while sparing adjacent healthy tissues. More often than not, the administration of radiotherapy is combined with other treatment options like; surgery, chemotherapy, targeted therapy, hormone therapy, etc. Either as neoadjuvant, adjuvant, or concurrent therapy.

What is Big Data

Radiotherapy treatment has attracted a great deal of research interest over the past decades, leading to significant technological advancement and complexity in the treatment workflow processes (3.1). These changes have generated a sea of rich digital information with a unique combination of patients' clinical demographics, images ("radiomics"), and biomarkers (genomics, proteomics, metabolomics) collected during the treatment period. This sea of patient information makes radiation oncology a fertile and perfect ground for learning and advancing care through Big Data opportunities (1).

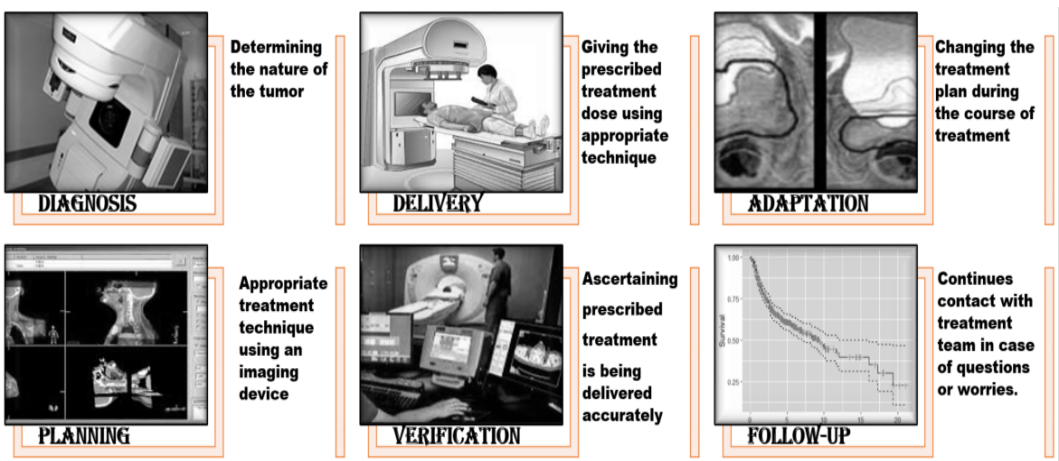

Figure 3.1. Radiotherapy workflow.

Big Data is not a new concept in radiation oncology, and it existed long before the term was coined in 2005. Records of Big Data application and the quest to analyse the available data to generate information that could shape our activities dates back more than 7,000 years. As with all new terminology, a diverse range of definitions has been proposed to describe Big

Data best. However, *Gartner* defines Big Data as:

> "*Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation* ".

As one of the most "hyped" terms in recent times, Big Data is no longer restricted to only the big three **V**'s (Volume, Velocity, and Variety) as defined above. As the amount of collected data from different sources continuously grow big and complex, with poor structure and incompleteness, a fourth **V** (veracity) is added to the mix.

Features of Big Data in radiation oncology

Volume
Big Data has to be "big" before everything else and in this context measured in terms of volume. As radiation therapy is used in about 50% of cancer patients producing up to 10GB of data in advanced Image-Guided Radiation Therapy (IGRT) in the developed world, the volume of data is indeed large.

Velocity
The technological advancements in the field of radiation oncology in the last decades have led to the development of radiation therapy machines, which are faster and more precise and produce images before and during treatment. These enhancements mean an increased velocity at which new data is being generated. This accelerated rate of data generation, together with the adoption of the electronic health records systems (EHRs), makes radiation therapy a natural and fertile ground for Big Data decision making in real-time (or near real-time).

Variety
With an increasing rate at which data is being generated (Velocity) leads to a large amount of data being collected (Volume), which brings in the third feature variety. This feature describes the vast diversity of data types in radiation therapy. From the radiation treatment standpoint alone, a patient can receive either external or internal radiation with a variety of modalities (MRLinac, particle therapy, etc.), each of these methods generates different forms of data. Combining them with that from other sectors (radiology, surgical oncology, administrative, insurance and social media) brings in all types of data formats (numbers, images, videos, text, or audio).

Veracity
Being able to trust a data source or quality (i.e., knowing who generated the data, how it was collected, when it was generated, and the purpose for collecting the data) is more pivotal than accessing the data when the intent is inference. Big Data in radiation therapy come from different institutions all around the world, and these institutes use various instruments and equipment to collect these data. Implying that data elements could be missing, biased, inconsistent, or not precise.

Figure 3.2 provides a graphic summary of Big Data structure.

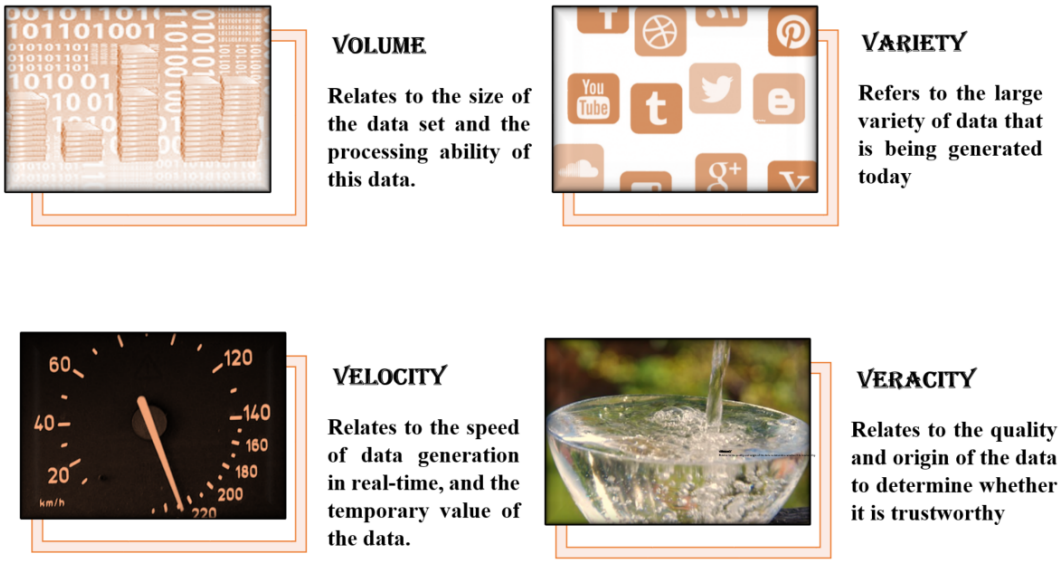

Figure 3.2. Big Data structure

With the sea of diverse data generated as a by-product of routine clinical workflow, coupled with the fact that radiation therapy is the most opted form of cancer treatment gives the field of radiation oncology a unique advantage in using Big Data to generate insights that will benefit millions of cancer patients undergoing radiation therapy worldwide.

The primary objective of Big Data in radiation oncology is to improve the quality of life after treatment. Hence, accumulating a vast amount of data which cannot generate insights that can guide our future actions is of no value. Investing in Big Data infrastructure merits a better understanding of the data generation source.

Big Data sources in Radiation Oncology

Big Data is gradually changing the field of radiation oncology, with the promise of improving the quality of care, decision-making, and assist in some of the patient care workflow tasks. However, these promises are based on the concept that a large amount of aggregated data will yield valuable insights that can help make reliable decisions. Hence, ensuring that the data collected is of high quality and integrity is very pivotal because achieving high-quality data is as relevant and pivotal as the question to be answered.

Radiation therapy process begins when an individual, usually referred by the clinical oncologist, first comes for consultation after the diagnostic phase is completed. The radiation oncologist discusses the treatment strategy and the clinical situation with the patient. After the consultation and assessment phase, the patient follows a series of stages

outlined in figure 3.1. All these stages of the radiotherapy workflow generate an enormous amount of data, which can be termed Big Data.

The data in radiation oncology can be classified into three categories. These categories include diagnostics and prognostics, treatment and symptoms management, and outcome measures. The goal is to use the diagnostic and prognostic factors for informed decision-making at the treatment stage for a better outcome.

Diagnostic phase

Whenever a person is diagnosed with cancer, an image scan and pathological tests are conducted to ascertain the stage of the disease and to verify if the tumor has metastasized to other parts of the body like the lymph nodes (glands). Some of the medical image modalities that produce detailed physiological and anatomical images of the body (2) with varying contrast are film x-rays, computed tomography (CT), ultrasound (US), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT), and positron emission tomography (PET) (Figure 3.3). Mostly the three 3D imaging modalities (MRI, CT, and PET) are used for diagnosing tumors(3).

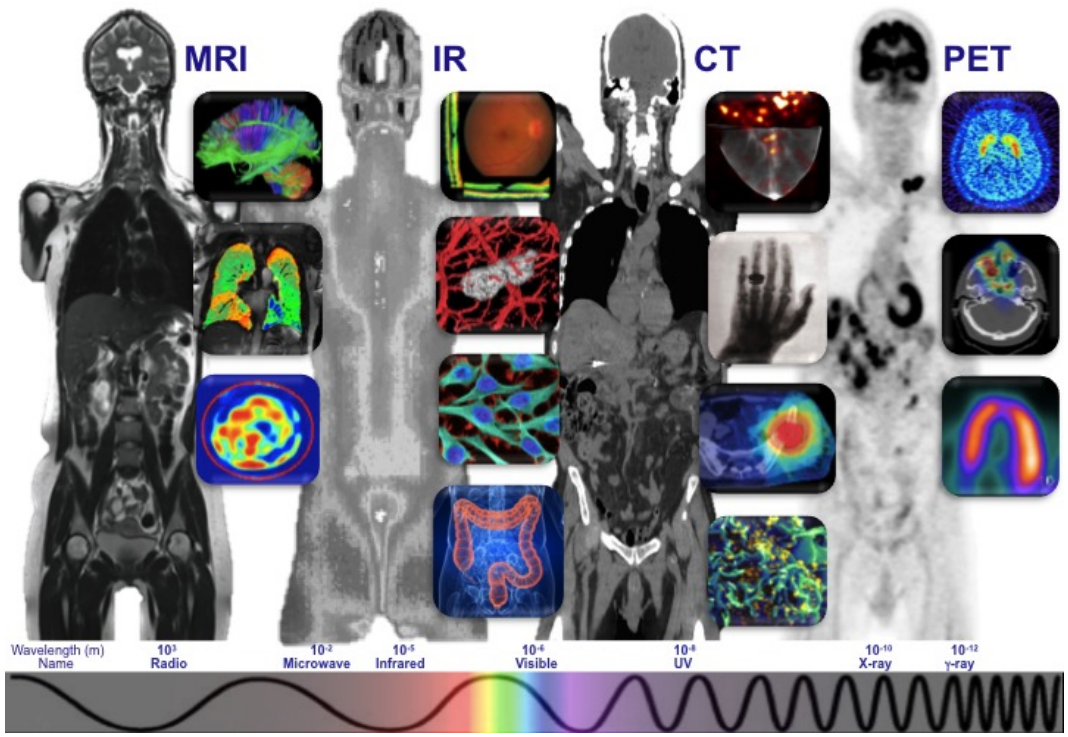

Figure 3.3. Image modalities used in clinics to visualize the human anatomy.
Source: Journal of Medical Imaging cover image, Vol. 2, No. 3 (2015).

With the technological advancement in the last decades, developers have developed tools which can convert these medical images into hundreds of thousands of (quantitative) features, identifying specific tumor patterns in the image pixels. These extracted image features can be categorised into four groups (shape, volume, texture, and intensity). Other feature categories like log and wavelet can be computed with some advanced image pre-processing.

Approximately 140 million patients were diagnosed with cancer before 2017 (4), with over 18.1 million new cases estimated in 2018 (5). Assuming that 50% of this population receives radiation treatment, this implies a minimum (one study per patient) of 79 million patient studies exist. These millions of images provide a potential wealth of patient data. Biopsies taken in the diagnostic phase lead to another huge and growing source of data, from pathology images and reports to biological data ranging from mutation to "panomics" data of normal and tumor tissue.

Treatment planning data

The planning of the appropriate treatment technique for a radiation oncology treatment is a team effort by oncologists, therapists, physicists, and dosimetrists. First, a treatment simulation session is conducted without actually radiating the patient where typically imaging scans (CT, MRI, PET) with the patient in treatment position are generated. Subsequently, the target volume or organs at risk are delineated; these Regions of Interests (ROI) are another data source. More data is created when the treatment plan is made, and the radiation dose is calculated, including derived measures such as the dose-volume histogram.

Treatment Verification and Delivery

Treatment verification using images (IGRT) and dosimetric verification before and during treatment (DGRT) produce data on the actual delivered dose and the quality of the therapy. Specific datasets might include cone-beam CT, kV images, megavoltage images, surface scans, ultrasound, and MRI. Furthermore, derived information is created, such as mismatches in position and subsequent shifts in table position or differences in planned versus delivered dose. A more recently available data source is the detailed logs in machine settings and sensor data (e.g., MLC positions). Finally, the treatment record is stored, containing process information such as when which beam was given.

Treatment Adaptation

Adaptive radiotherapy is the changing of the radiation treatment plan of a patient during treatment to account for unplanned changes (6). Typical reasons to adapt are tumor shrinkage or patients losing weight during the course of treatment. These patient-tailored adaptations of therapy are aimed at improving tumor outcomes and reducing toxicity. New

images, plans, regions-of-interest, plans, and dose distribution are created, all potentially interesting data sources.

Follow-up

Once a patient has completed the course of radiation therapy, several regular follow-up visits are scheduled by the radiation oncologist or referring doctor. These visits are to examine and check the results of the treatment and adverse effects and often include repeating tests used in the diagnostic phase such as imaging, lab tests, physical examinations, etc. All these processes generate additional patient data. Of particular interest are doctor and patient-reported outcomes, which are typically standardized measures of tumor control, early and late toxicity, and quality of life.

Other treatments

Radiotherapy is rarely given as monotherapy. Depending on the tumor stage, surgical resection might be indicated and comes with its own data sources such as pathology, anesthesia, surgical reports, etc. Patients might also receive chemotherapy as a (neo)adjuvant therapy to the local treatment, depending on the invasive nature of the tumor (7–9).

The agent given, the number of cycles, and acute toxicities are generally interesting data to combine with radiotherapy data. Other treatment modalities include but not limited to bone marrow transplant, immunotherapy, hormone therapy, targeted drug therapy, cryoablation, radiofrequency ablation, and hypothermia. These treatment options might be administered in combination with each other and with radiotherapy – all producing data which might be relevant.

Non-medical data

Although a focus on medical information is understandable, it has been realized that non-medical information such as social-economic status, social network, education, lifestyle, and environment might play an essential role in cancer detection, treatment, and outcomes, as shown in Figure 3.4.

Besides factors that may determine outcomes, surrogate measures to determine outcomes more efficiently, such as using insurance claims for tumor recurrence and wearable and mobile devices for quality of life, are interesting data sources to consider. Including such non-medical data sources is rarely done but can be expected to occur more often in the future.

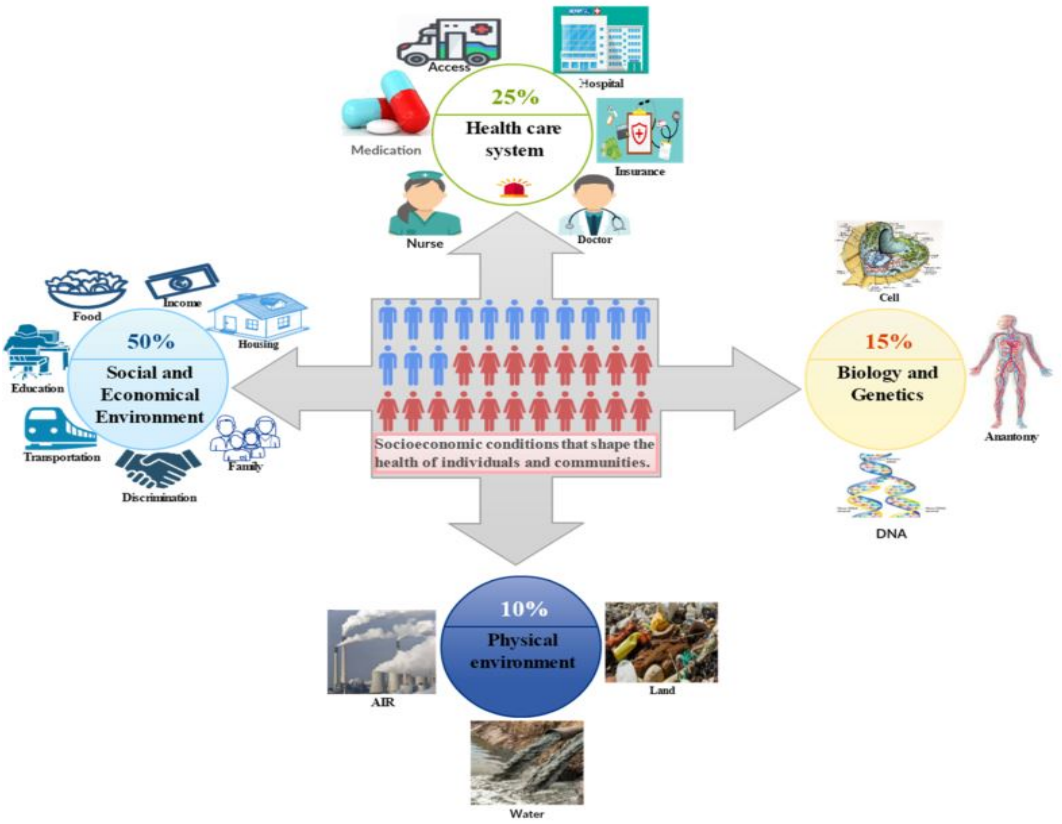

Figure 3.4. Socio-economic conditions that shape the health of individuals and communities.

In conclusion, data in radiation oncology is firmly in the realm of "Big Data." With the availability in many departments of electronic health records systems (EHR), oncology information systems (OIS), and picture archiving and communication systems (PACS), data about radiation oncology patients are becoming more and more available and with higher quality. From calculations by experts, the annual collection of these data for a patient will add 4 and 76 megabytes of text and imaging data to the EHRs, respectively (10, 11). As estimated in subsection 3, assuming these 79 million radiotherapy-treated patients survive for one year, then at the very least 5.886 petabyte (4 + 76 megabytes x 79 million patients) of radiation oncology data exist. As discussed before, these data in radiation oncology are considered as *"Big Data"* because they can be classified into:

Volume: The extensive use of imaging and quantitative analyses.

Velocity: Growing adoption of image-guided and adaptive radiotherapy.

Variety: The use of different imaging and treatment modalities and structured an unstructured information

Veracity: Many different data sources with varying quality

Recently, data has been considered as the new oil due to the way it is used to power much of the transformative technologies we see today. Although valuable, data needs processing, just as oil needs refining for it to be useful. The radiation oncology field is actively interested in benefiting from this Big Data process output from two perspectives. First, to automate and improve the treatment processes to treat patients in the most cost-effective manner possible while ensuring a high quality of care. Second, for better cancer care by making sure that patients have the best treatment outcome in terms of patient priorities and precision medicine (12).

Big Data for better processes and automation

There have been a handful of novel applications of Big Data in radiation oncology recently. Automating some parts of the radiation therapy workflow has produced a lot of benefits, from reducing operational costs to improving operational efficiency.

Auto-contouring

The accurate contour delineation of the target region and organs at risk (OAR) is an essential part of treatment planning (13). Several diagnostics images as mentioned in section 3 can be used in the tumor and OAR structures delineation process. However, manual-contouring is time consuming with high level interhuman variation in the quality of contours due to skills level (14). Numerous robust and efficient delineation tools have been developed with the dawn of the Big Data era in radiation therapy to help automate and speed these delineation process (13). Some of these Big Data applications tools includes auto-segmenting vital organs of the male pelvic anatomy (prostate, bladder, rectum) on CT-based images (15). Automatic segmentation of brain organs at risk (16, 17).

Quality assurance

Quality assurance (QA) in radiation therapy are those procedures that ensure consistency of the prescription and the safe fulfilment of that prescription in terms of dose to the target volume, together with minimal dose to healthy tissue, minimal exposure of personnel, and adequate patient monitoring aimed at determining the result of treatment (18). It should be noted that QA here spans the entire radiation therapy workflow process to limit every source of errors and ensure a high standard of radiation treatment. Big Data QA methods have been developed to detect, minimize, and prevent these errors and anomalies in radiation therapy (19, 20).

Image reconstruction

We have seen the different image modalities in subsection 3. Here, Big Data for image reconstruction can be placed into two categories.

> *The first used case is the mapping of and image, from one imaging modality format to another (CT image from MRI).* This Big Data application to transition from one image modality to another is beneficial to bridge the limitation of one image modality for treatment. CT images are used for dose calculation in the treatment process, but they have inferior image contrast and possible side effect due to radiation. On the other hand, MRI, which is safe (no radiation involve) with enhanced image contrast, is very expensive.
>
> *The second use is the reconstruction of a high-quality image from a low-quality images (20).*

Big Data for better outcomes

Our understanding of the human body has improved tremendously in the last decades due to the advancement in medical technology. These technological advancements have empowered our abilities to extract phenotypic features of tumors from medical images, study tumors relations to their surrounding at a cellular level (21, 22), and apply gene silencing for therapeutic reasons (23). Big Data can make use of this large volume of medical information from individual patient data to identify novel risks or therapeutic options that can then be applied at the individual patient level to improve outcomes. However, to make use of these new Big Data techniques in radiation therapy, rigorous testing in real-life scenarios and a lot of collaboration will be required.

Decision support

Decision-making is part of every daily human life. **Myles Munroe** said *"Our life is the sum of all the decisions we make daily"*. Making an optimal decision is not easy (21). In recent times, decision-making has been likened to problem-solving, making decision-making be perceived as a problem-solving action. Therefore a system that can aid people to make the optimal decision is very pivotal. Systems that are capable of providing not just information, but can participate in simple decision-making activities of an organization, are known as Decision Support Systems (21). A decision support system (DSS) is a computer-based application that collects, organizes, and analyzes data to facilitate quality decision-making for the management, operations, and planning of the organization. DSS can be very handy in complicated circumstances or rapidly changing situations where anticipating or determining the future outcome is very difficult (24). Complex biological systems like the human body are suitable areas where information technology can be applied to reduce error and improve decision quality.

One of the areas where the application of information technology is very pivotal is in clinical decision support systems (CDSS) (21). A CDSS is a health information technology system that is designed to assist physicians and other health professionals with their clinical decision-making tasks by integrating different sources of health care information such as electronic health records (EHR), laboratory test results, etc. (25). In recent years, CDSS have become an essential topic for Big Data in radiation oncology with numerous domain experts (radiologist and doctors) working together to build these Big Data CDSS tools. Diverse Big Data technologies have been applied in radiation therapy to help radiologists and physicians in their decision-making process (20, 25) like diagnostic support systems which can detect lung nodules in thoracic CT scans of a patient (26), brain lesion on an MRI (27) or micro-calcification breast masses in mammograms (28)

Precision medicine

During treatment development, the target is always for an "average" patient, ignoring the reality of patients not being alike. The omission of this potential patient variability during the development stage flows down to the treatment stage. In a therapeutic setting, patients react differently to treatment regimes. Some patients do not respond to the treatment at all, some patients' condition becomes worse or deteriorates, and others recover quickly under the same treatment condition (29). This variation in treatment response may depend on observable factors such as genetic polymorphisms, age, sex, or tumor imaging characteristics which opens up the option to personalize treatment or precision medicine.

In the field of oncology, precision medicine can be defined as a medical practice where patient-specific tumor and normal tissue information is used to aid in the treatment planning, diagnosis, prognosis, and ascertain treatment response, in other words, identifying which treatment approach will be more useful for a patient based on genetic, environmental, and lifestyle factors (Figure 3.5). Personalized medicine is an older term conveying the same meaning as "precision medicine". However, the term is sometimes misinterpreted as implying that a unique treatment or prevention can be developed for each individual (30–33).

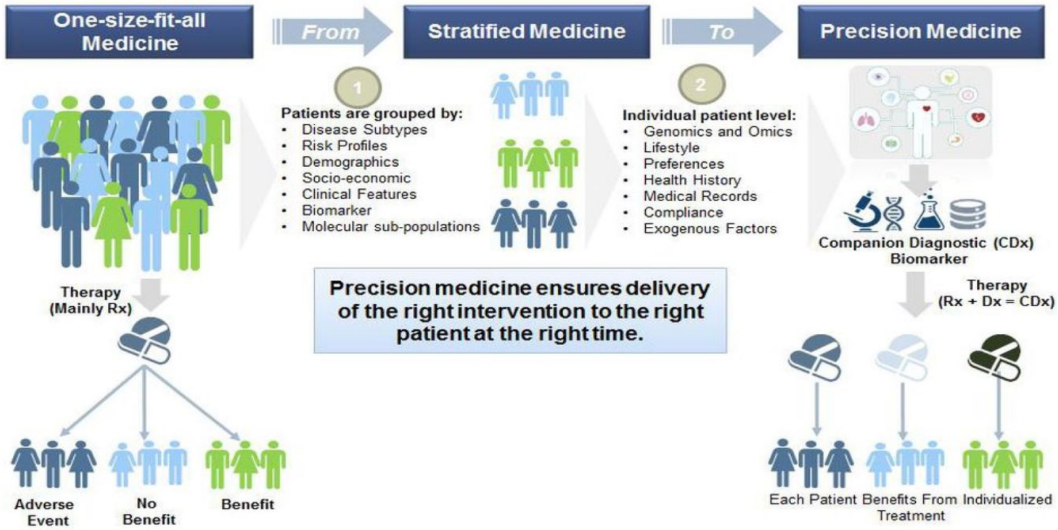

Figure 3.5. Distinction of precision medicine treatment form other treatment forms.
Source: Frost & Sullivan-New Paradigm Shift in Treatment

Precision medicine which aims at classifying patients into sub-populations based on their disease susceptibility or treatment response is thought to be the best treatment option because it removes the "one size fits all" treatment paradigm allowing therapeutic interventions to be channeled to patients with the most benefits. Hence reducing the cost of treatment, improve quality of life in terms of treatment burden or side effects, and waste of time for those without a treatment benefit (33). Using Big Data in radiation oncology might make the dream of precision medicine a reality by being able to predict patient-specific outcomes of future patients (22).

Barriers of using Big Data applications in clinical

Big Data in radiation oncology, like in any other field, uses the same concept of looking for trends and associations from large volumes of individual patient medical information such as found in electronic health records systems (EHRS), images, insurance claims, social media groups, and wearable and mobile devices to identify novel risks or therapeutic options that can then be applied at the individual level to improve outcomes of future patients. So, the million-dollar question is

"Why is Big Data in radiation oncology not widely accepted like in other industries?"

Google, for example, uses the same concept to provide accurate and personalized real-time information to its many users. Here are some possible reasons why.

Data

The very first evident limitation is data. Even though we are in the era of Big Data, Big Data in radiation oncology is not yet big enough as compared to the likes of Google, Netflix, Facebook, Ali-express, etc, which have free and open access to an ocean of centralized data willingly provided by the customer. This limitation is chiefly because Big Data in radiation oncology is scattered across multiple hospitals and healthcare centers with no central sharing to allow the velocity and volume of data required to exploit Big Data methods properly.

Privacy

The second limitation is privacy. Medical record, which represents deep individual patients' private and personal information is heavily guarded in their respective silos and not readily shareable without violating patients' privacy rights.

Technical Know-How

As complex as the human body is, so is the medical data extracted. Therefore, a well-developed infrastructure and technicians with medical and data management knowledge are required to process, manage and maintain these medical data from different sources into a readily usable form. However, persons with this unique combination of skill sets are limited and hard to find (34).

Reputation

Despite the fact that we are in the era of Big Data and everybody is anxious to jump into this Big Data train to benefit from its many opportunities, many hospitals and healthcare organization shy away from joining this Big Data train because they cannot accurately value their current data. Sharing their data might jeopardize their reputation as it allows their performance to be compared with others (e.g. higher mortality and recurrence than others). Furthermore, these data can be used to build Big Data techniques for which they have sole monopoly over. This implies competing institutes can deliver these techniques as well, hence creating unwanted competitors.

So, regardless of the technological advancements in many areas of radiation oncology, there still exist significant barriers in accessing patient data. Circumventing these barriers and making Big Data in radiation oncology solutions work as intended will entail some effort. First, a proper ontology with formal and well-defined terminologies (types, properties, and interrelationships) within the radiation oncology domain (Radiation Oncology Ontology (ROO), National Cancer Institute Thesaurus (NCIT) ontology and International Classification of Diseases (ICD) ontology). These ontologies are organized in a formal logical format and standardized in both human-readable and machine process-able format then shared with the many participants (35, 36). Each participating institution is in charge of maintaining the ontology, de-identifying their data and getting prior informed consent for the use of these de-identified data.

Experience of bringing to practice

The main reason why new Big Data products are not readily accepted into practice is that it is not trivial to figure out where to insert these technologies into the existing working workflow (Figure 3.1). Introducing new data-driven techniques into the current radiation therapy workflow will lead to a significant disruption in the workflow causing specific professions and radiation oncology as a whole to undergo changes that may be resisted.

Another reason is trust. These new data-driven technologies need to go via long periods of testing before they can be trusted, especially techniques that replace a human action in some way like diagnosing if a patient has a tumor from an image. But, technologies that automate some parts of the workflows or reduce human workload and speed up the radiation therapy process like auto contouring might transition rapidly to daily use. Similarly, techniques not previously possible due to lack of Big Data infrastructures like auto-contouring are likely to gain early adoption. In essence, Big Data technologies that are not trivial to use or will require interrupting the normal radiation therapy process will meet some resistance transitioning to practice.

On the other hand, Big Data technologies which are entirely built (hardware and software incorporated) or requires just a push of the button on a computer screen to activate the program in the background are far more likely to be adopted especially if they can solve time-consuming tasks like auto-contouring. As previously mentioned, technologies that can automate the radiation therapy workflow or are capable of flagging suspicious regions on complex images and present a structured quantitative and visual report requiring just a simple visualization check or a mouse click validation by the users will see greater success transitioning into the clinics.

The future of Big Data in RO with FAIR principles

A significant number of patient data and databases have been established over the past decades worldwide. However, this data is not often in the same place but scattered over multiple hospitals and care centers around the globe with very little knowledge of their existence, accessibility, collection protocol or generation procedure. This limits the potential of radiation oncology Big Data in providing relevant and critical insights, that can then be beneficial to future patients. At the same time, accumulating an extensive amount of data is of no value if it cannot be used to generate future insights which can shape our thinking and actions. To benefit from Big Data in radiation therapy, these databases should follow the FAIR guiding principles (37). In other words, they should be Findable, Accessible, Interoperable, and Reusable (Figure 3.6) while still respecting the patient's privacy and medical confidentiality.

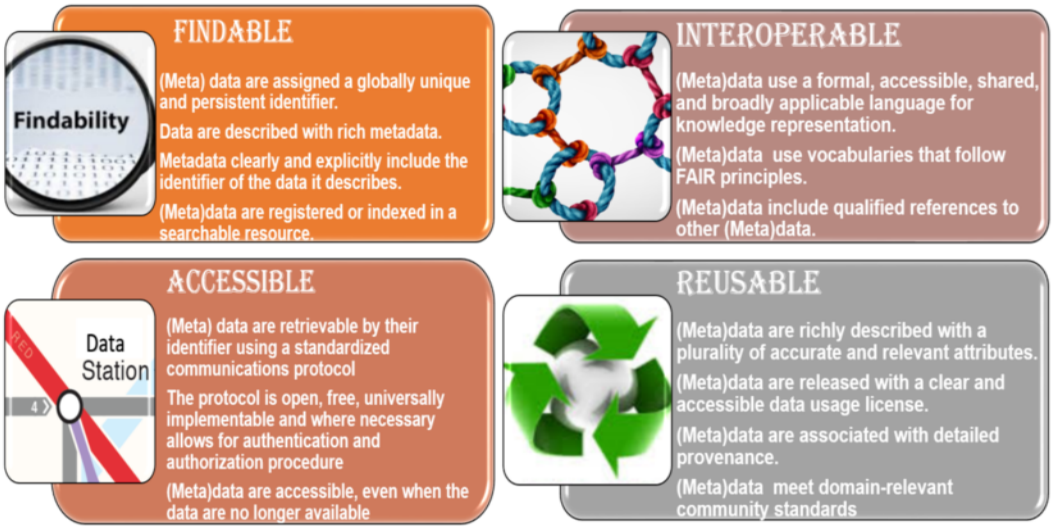

Figure 3.6. FAIR Guiding Principles for scientific data management and stewardship

Findable
The data should be easily identified and found for both humans and computers, with the metadata to facilitate the search for the specific datasets.

Accessible
The data should be stored for as long as possible so that they can be accessed easily or downloaded with well-defined accessing conditions, be it at metadata level or the actual data.

Interoperable
The data should be ready to be combined with other datasets either by humans or computers, without any ambiguous meanings of the data terms and values.

Reusable
The data should be prepared for future research use and further processed with computational methods. This requires adequate information on the data generation step and processing with an appropriate license.

Summary

The era of Big Data in radiation oncology is already here, with numerous challenges as well as opportunities in the radiation oncology domain. There has been a limited number of novel applications of Big Data in radiation oncology recently, with useful and appropriate tools to assist in decision-making for clinicians, personalize treatment strategy for an individual patient, and improve patient safety and treatment efficacy. Automating parts of the radiation therapy workflow, auto-contouring are expected to be the first of many Big Data application

Bibliography

[1] Stanley H Benedict, Issam El Naqa, and Eric E Klein. Introduction to big data in radiation oncology: Exploring opportunities for research, quality assessment, and clinical care. *International Journal of Radiation Oncology• Biology• Physics*, 95(3):871–872, 2016.

[2] Giger Maryellen L and Tornai Martin P. The spectrum of medical imaging in the 2015 international year of light. *Journal of Medical Imaging*, 2(3):1 – 2 – 2, 2015. doi: 10.1117/1.JMI.2.3.030101. URL https://doi.org/10.1117/1.JMI.2.3.030101.

[3] J Deng. Big data in radiation oncology: challenges and opportunities. *Cancer Sci Res Open Access*, 1(2):1–2, 2014.

[4] Tim Lustberg, Johan van Soest, Arthur Jochems, Timo Deist, Yvonka van Wijk, Sean Walsh, Philippe Lambin, and Andre Dekker. Big data in radiation therapy: challenges and opportunities. *The British journal of radiology*, 90(1069):20160689, 2017.

[5] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424, 2018.

[6] Shrikant Balasaheb Mali. Adaptive radiotherapy for head neck cancer. *Journal of maxillofacial and oral surgery*, 15(4):549–554, 2016.

[7] Scott Kopetz, George J Chang, Michael J Overman, Cathy Eng, Daniel J Sargent, David W Larson, Axel Grothey, Jean-Nicolas Vauthey, David M Nagorney, and Robert R McWilliams. Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. *Journal of clinical oncology*, 27(22):3677, 2009.

[8] Jean-François Bosset, Laurence Collette, Gilles Calais, Laurent Mineur, Philippe Maingon, Ljiljana Radosevic-Jelic, Alain Daban, Etienne Bardet, Alexander Beny, and Jean-Claude Ollier. Chemotherapy with preoperative radiotherapy in rectal cancer. *New England Journal of Medicine*, 355(11):1114–1123, 2006.

[9] Robert Rosenberg, Ken Herrmann, Ralf Gertler, Beat Künzli, Markus Essler, Florian Lordick, Karen Becker, Tibor Schuster, Hans Geinitz, Matthias Maak, et al. The predictive value of metabolic response to preoperative radiochemotherapy in locally advanced rectal cancer measured by pet/ct. *International journal of colorectal disease*, 24 (2):191, 2009.

[10] Brown Nathan. *Healthcare Data Growth: An Exponential Problem*, 2015. https://www.nextech.com/blog/healthcare-data-growth-an-exponential-problem [Accessed: 2019-07-20].

[11] Halamka John. *The Cost of Storing Patient Records*, 2011. http://geekdoctor.blogspot.com/2011/04/cost-of-storing-patient-records.html [Accessed: 2019-07-20].

[12] Akuli B Osong, Andre Dekker, and Johan van Soest. Big data for better cancer care, 2019.

[13] Xingyu Wu, Jayaram K Udupa, Yubing Tong, Dewey Odhner, Gargi V Pednekar, Charles B Simone, David McLaughlin, Chavanon Apinorasethkul, John Lukens, Dimitris Mihailidis, et al. Auto-contouring via automatic anatomy recognition of organs at risk in head and neck cancer on ct images. In *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*.

[14] Shalini K Vinod, Michael G Jameson, Myo Min, and Lois C Holloway. Uncertainties in volume delineation in radiation oncology: a systematic review and recommendations for future studies. *Radiotherapy and Oncology*, 121(2):169–179, 2016.

[15] Stuart Greenham, Jenna Dean, Cheuk Kuen Kenneth Fu, Joanne Goman, Jeremy Mulligan, Deanna Tune, David Sampson, Justin Westhuyzen, and Michael McKay. Evaluation of atlas-based auto-segmentation software in prostate cancer patients. *Journal of medical radiation sciences*, 61(3):151–158, 2014.

[16] Aurélie Isambert, Frédéric Dhermain, François Bidault, Olivier Commowick, Pierre-Yves Bondiau, Grégoire Malandain, and Dimitri Lefkopoulos. Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. *Radiotherapy and oncology*, 87(1):93–99, 2008.

[17] Manuel Conson, Laura Cella, Roberto Pacelli, Marco Comerci, Raffaele Liuzzi, Marco Salvatore, and Mario Quarantelli. Automated delineation of brain structures in patients undergoing radiotherapy for primary brain tumors: From atlas to dose–volume histograms. *Radiotherapy and Oncology*, 112(3):326–331, 2014.

[18] DAVID I Thwaites, BJ Mijnheer, and JOHN A Mills. Quality assurance of external beam radiotherapy. *Radiation oncology physics: a handbook for teachers and students*, pages 470–50, 2005.

[19] Todd R McNutt, Kevin L Moore, Binbin Wu, and Jean L Wright. Use of big data for quality assurance in radiation therapy. In *Seminars in radiation oncology*, volume 29, pages 326–332. 2019.

[20] Alexander FI Osman. Radiation oncology in the era of big data and machine learning for precision medicine. In *Machine Learning in Medicine and Biology*. IntechOpen, 2019.

[21] Amir Mohammad Shahsavarani, Esfandiar Azad Marz Abadi, Maryam Hakimi Kalkhoran, Saeideh Jafari, and Shirin Qaranli. Clinical decision support systems (cdsss): state of the art review of literature. *International Journal of Medical Reviews*, 2(4):299–308, 2015.

[22] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2015.

[23] Behzad Mansoori, Siamak Sandoghchian Shotorbani, and Behzad Baradaran. Rna interference and its role in cancer therapy. *Advanced pharmaceutical bulletin*, 4(4):313, 2014.

[24] Alexander Boukhanovsky and Marian Bubak. High performance computations for decision support in critical situations: Introduction to the third workshop on urgent computing. *Procedia Computer Science*, 29:1644–1645, 2014.

[25] Tim Bezemer, Mark CH de Groot, Enja Blasse, Maarten J ten Berg, Teus H Kappen, Annelien L Bredenoord, Wouter W van Solinge, Imo E Hoefer, and Saskia Haitjema. A human (e) factor in clinical decision support systems. *Journal of medical Internet research*, 21(3):e11732, 2019.

[26] Kenji Suzuki, Samuel G Armato III, Feng Li, Shusuke Sone, and Kunio Doi. Massive training artificial neural network (mtann) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography. *Medical physics*, 30 (7):1602–1617, 2003.

[27] Heang-Ping Chan, Shih-Chung B Lo, Berkman Sahiner, Kwok Leung Lam, and Mark A Helvie. Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network. *Medical Physics*, 22(10):1555–1567, 1995.

[28] Mina Rezaei, Haojin Yang, and Christoph Meinel. Deep neural network with l2-norm unit for brain lesions detection. In *International Conference on Neural Information Processing*, pages 798–807. 2017.

[29] Nicholas J Schork. Personalized medicine: time for one-person trials. *Nature News*, 520 (7549):609, 2015.

[30] Ewout W Steyerberg and Yvonne Vergouwe. Towards better clinical prediction models: seven steps for development and an abcd for validation. *European heart journal*, 35(29): 1925–1931, 2014.

[31] Philippe Lambin, Steven F Petit, Hugo JWL Aerts, Wouter JC van Elmpt, Cary JG Oberije, Maud HW Starmans, Ruud GPM van Stiphout, Guus AMS van Dongen, Kristoff Muylle, Patrick Flamen, et al. The estro breur lecture 2009. from population to voxel-based radiotherapy: exploiting intra-tumour and intra-organ heterogeneity for advanced treatment of non-small cell lung cancer. *Radiotherapy and oncology*, 96(2): 145–152, 2010.

[32] David C Whitcomb. What is personalized medicine and what should it replace? *Nature Reviews Gastroenterology and Hepatology*, 9(7):418, 2012.

[33] Ronald C Chen, Peter E Gabriel, Brian D Kavanagh, and Todd R McNutt. How will big data impact clinical decision making and precision medicine in radiation therapy? *International Journal of Radiation Oncology• Biology• Physics*, 95(3):880–884, 2016.

[34] Emilie Baro, Samuel Degoul, Régis Beuscart, and Emmanuel Chazard. Toward a literature-driven definition of big data in healthcare. *BioMed research international*, 2015, 2015.

[35] Elisa Meldolesi, Johan van Soest, Nicola Dinapoli, Andre Dekker, Andrea Damiani, Maria Antonietta Gambacorta, and Vincenzo Valentini. An umbrella protocol for standardized data collection (sdc) in rectal cancer: a prospective uniform naming and procedure convention to support personalized medicine. *Radiotherapy and Oncology*, 112(1):59–62, 2014.

[36] Jean-Emmanuel Bibault, Eric Zapletal, Bastien Rance, Philippe Giraud, and Anita Burgun. Labeling for big data in radiation oncology: the radiation oncology structures ontology. *PloS one*, 13(1):e0191263, 2018.

[37] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3, 2016.

Part III
Regression

CHAPTER 4

PREDICTION SCORING SYSTEM FOR CERVICAL CANCER PATIENTS

Authors:
Youn Ji Kim, Young Saing Kim, Jin Woo Shin, ***Biche Osong****, Seok Ho Lee*

Adapted from*: The International Journal of Gynecological Cancer (IJGC)*
DOI*:* ***10.1136/ijgc-2019-001050***

Abstract

Purpose
Scoring system based on clinicohematologic parameters in cervical cancer patients receiving chemoradiation has not been reported to date. The aim of this study was to determine the prognostic value of clinicohematologic parameters in patients with cervical cancer undergoing chemoradiation and to develop a prediction scoring system based on these results.

Summary of background
Cervical cancer is the seventh most common cancer and the third leading cause of cancer-related death. Although most of these deaths occur in areas where healthcare services are indigent and limited, numerous factors are associated with these death rates.

Patients and Methods
A total of 107 patients who received definitive chemoradiation for cervical cancer were enrolled in this study. The clinical data and hematologic parameters were retrospectively reviewed, and their prognostic value in predicting survival was analyzed. The neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and lymphocyte-to-monocyte ratio (LMR) and the changes in these hematologic parameters ($\triangle$NLR, $\triangle$PLR, and $\triangle$LMR) between pre- and post-treatment were calculated to determine the specific value of these parameters for predicting patient survival.

Results
The median follow-up time was 39.9 (range 2.7 - 114.6) months. The 3-year overall survival rate and progression-free survival rate were 80.9% (95% CI: 72.7 - 90.0) and 53.4% (95% CI: 44.1 - 64.8), respectively. The median progression-free survival was 67.5 months and the median overall survival was not reached. According to multivariable analysis, a $\triangle NLR \geq 0$ was significantly associated with decreased progression-free survival (HR = 2.91, 95% CI 1.43 - 5.94) and overall survival (HR = 3.13, 95% CI: 1.18 - 8.27). In addition, age (age <58.5 years; progression-free survival: HR = 2.55, 95% CI 1.38 - 4.70; overall survival: HR = 4.49, 95% CI: 1.78 - 11.33) and the International Federation of Gynecology and Obstetrics (FIGO) stage (III-IV progression free survival: HR = 2.49, 95% CI: 1.40 - 4.43; overall survival: HR = 3.02, 95% CI: 1.32 - 6.90) were identified as predictors of poor survival.

Conclusion
Both the age and FIGO stage, as clinical parameters, and the $\triangle$NLR, as a hematologic parameter, were independent prognostic factors for survival for cervical cancer patients treated with chemoradiation. Based on these results, we developed a risk score-based classification system for predicting survival.

Introduction

Cervical cancer is the seventh most common cancer and the third leading cause of cancer-related death (1). Currently, a comprehensive approach including prevention, enhanced screening, and early diagnosis and treatment has reduced the high incidence and mortality rate associated with cervical cancer. Nevertheless, the prevalence of human papillomavirus remains high in Korea (2), and the incidence of cervical cancer in Korea is higher than that in the USA and European countries (3).

In Korea, the incidence of cervical cancer decreased from an age-standardized rate of 16.3 per 100 000 in 1999 to 9.1 per 100 000 in 2015. However, cervical cancer was still the most common gynecologic cancer during that period (4).

Parametrial invasion, lymph node metastasis, large-sized tumors, and advanced clinical stage have been reported as poor prognosis indicators (5). In cervical cancer, an increased neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) have also been reported to be related to a poor prognosis (6, 7). Furthermore, patients with a decreased lymphocyte-to-monocyte ratio (LMR) have been reported to have a poor prognosis (8).

The change in the NLR was a significant prognostic predictor for advanced pancreatic cancer patients who received chemotherapy (9). However, the prognostic significance of the change in the NLR in patients with cervical cancer who are receiving chemoradiation has yet to be determined. For the decision-making process, the prediction scoring system has already been shown to be an effective strategy for cancer patients receiving further treatment (10).

A scoring system based on clinico-hematological parameters in cervical cancer patients receiving chemoradiation has not been reported to the best of our knowledge. Thus, this study aimed to determine the predictive value of pre-treatment clinico-hematological parameters and changes in these hematological parameters in patients with cervical cancer treated with chemoradiation and to develop a prediction scoring system based on the results.

Materials and methods

Patient Selection

A total of 144 patients with histologically verified uterine cervical cancer received radiation therapy at our institution from January 2009 to December 2016. The patients in this study were staged according to the International Federation of Gynecology and Obstetrics (FIGO) 2009 staging system. The lymph node involvement for all patients was determined based on clinical imaging methods such as computed tomography (CT) or magnetic resonance imaging (MRI).

The inclusion criteria for this study were:

- Patients who were 18 years and older,
- Patients with FIGO stage[1] IB1, IIA, IIIA, IIIB, or IVA,
- Histopathologically verified squamous cell carcinoma[2],
- Patients without distant metastasis except para-aortic lymph node metastasis,
- Patient receiving definitive chemoradiation, including brachytherapy.

The exclusion criteria consisted of:

- Patients with incomplete information of interest[3],
- Incomplete radiation therapy course,
- Patient who received radiation therapy as a palliative treatment.

The patients underwent routine procedures, including physical examinations, laboratory tests, chest X-rays, and CT or MRI. The clinical data and CBC counts were retrospectively reviewed. The following variables were analyzed: age, FIGO stage, pathology, and lymph node involvement. The hematologic tests were performed before and after the definitive radiation therapy. The complete blood counts (CBC) included the hemoglobin levels, hematocrit, white blood cell counts, and platelet counts. The pre-treatment NLR (absolute neutrophil count/absolute lymphocyte count), PLR (absolute platelet count/absolute lymphocyte count), LMR (absolute lymphocyte count/absolute monocyte count), $\triangle$NLR (post-treatment NLR–pre-treatment NLR), $\triangle$PLR (post-treatment PLR–pre-treatment PLR), and $\triangle$LMR (post-treatment LMR–pre-treatment LMR) were calculated accordingly.

[1] Based on the 2009 staging system
[2] Adenocarcinoma, or Adenosquamous cell carcinoma
[3] Including laboratory data such as complete blood counts (CBC) performed before or after treatment

This study was approved by the Institutional Review Board (IRB) and performed following the international (Declaration of Helsinki) ethical standards. The study was exempted from informed consent due to the retrospective nature of this research and the fact that the analysis used anonymous clinical data.

Treatment

All patients received radiation therapy with a total dose in the range of 70.4 - 103.6 Gy. All patients received external radiation therapy using 10 or 15 MV photons in the pelvis using a four-field technique. A total external radiation therapy dose of 50.4 – 75.6 Gy was delivered in daily fractions of 1.8 – 2 Gy 5 days a week. After a median external radiation therapy dose of 54 Gy was given, brachytherapy was started using a microselectron high dose rate (Nucletron, The Netherlands) with an Ir-192 source. A total brachytherapy dose of 24 - 40 Gy was delivered twice a week with a 3 - 5 Gy fraction size. Patients who received chemotherapy were treated every week with 40 - 60 mg/m^2 cisplatin concurrently with external radiation therapy. In this study, no patient received consolidation chemotherapy after chemoradiation. Figure 4.1 shows the chemoradiation scheme used to treat cervical cancer. Treatment response was the primary endpoint. It was assessed by a clinical examination, including a gynecologic exam with a Pap smear, a CT or MRI scan from the chest to the pelvis, and positron emission tomography-computed tomography (PET/CT).

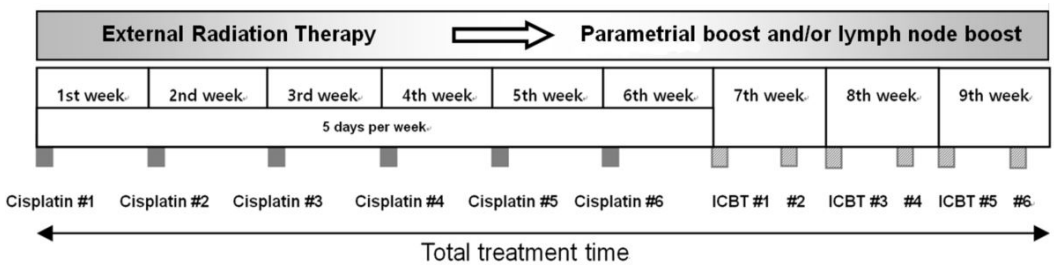

Figure 4.1. *Treatment scheme of a cervical cancer radiation therapy course incorporating external beam radiation therapy (EBRT) with concurrent weekly cisplatin, six fractions of intracavitary brachytherapy (ICBT), and a parametrial and/or lymph node boost.*

Statistical Analyses

The continuous variables were categorized based on the mean. The cut-off values for △NLR, △PLR, and △LMR were determined based on a past study (11). Progression-free survival was defined as the time interval between the initiation of treatment and the date of the first progression or death from any cause. Overall survival was defined as the time interval between the initiation of treatment and the date of death due to any cause or the date of the last contact. Lost to follow-up was defined as the incomplete ascertainment of the primary treatment outcome.

The Kaplan–Meier method was used to visualize patient survival over time, and the log-rank test for comparing survival differences between groups. The Cox proportional hazards regression model was used to determine the survival hazard ratios. According to the univariate analysis, the variables with p values < 0.3 were selected for multivariable analysis. For the development of the prediction scoring system, the β-coefficient-based scoring system was used (12). The β-coefficients were multiplied by 10 and rounded to the nearest integer. We assessed the model performance by calculating the concordance index and calibration plot. All p values were two-sided and were considered statistically significant if they were < 0.3. All statistical analyses were performed using SPSS (13) and R (14) software.

Results

Among the 144 patients who received radiation therapy at the Gil medical center and considered for this study, we excluded 11 patients who did not have enough laboratory test information, 3 patients who did not complete their scheduled treatment course, 2 patients treated for palliative reasons, 10 patients who did not receive brachytherapy, and 11 patients did not receive chemotherapy. A total of 107 patients who meet the eligibility criteria were included in the final analysis. In this study, the number of patients lost to follow-up was 11 (10.3%).

The patient characteristics are presented in Table 4.1. More than half of the patients had FIGO stage I and II cancer (I-II, 66.4%), and 67 patients (62.6%) had pelvic lymph node involvement. During external radiation therapy, most patients (n=93, 86.9%) who received chemoradiation were treated with cisplatin-based chemotherapy. The remaining 14 patients were treated with carboplatin. The median follow-up duration was 39.9 (2.7 - 114.6) months, and the 3-year overall survival rate and progression-free survival rate were 80.9 (72.7 - 90.0) and 53.4 (44.1 - 64.8), respectively. The total number of events in overall survival is 23 events, and the total number of recurrences is 47 recurrences. The median progression-free survival was 67.5 months, and the median overall survival was not reached.

The cut-off values for patient age, hemoglobin, hematocrit levels, white blood cell counts, platelet counts, NLR, PLR, and LMR, was 58.5 years, 11.45 g/dL, 34.6%, $6.98x103/\mu L$, $265.5x103/\mu L$, 2.33, 136.57, and 4.17, respectively. The cut-off values for the $\triangle$NLR, $\triangle$PLR, and $\triangle$LMR were determined to be zero (0) for survival (11).

Table 4.1. Summary of patients characteristics of the studies.

Variable	Levels	n (%)[range]	Variable	Levels	n (%)[range]
Patients	N	107 (100%)	Follow-up	Median	39.9 [02.7–114.6]
Age (years)	Median	55 [26–84]	Histology	SCC	95 (88.8)
	< 58.5	42 (39.3)		AC	08 (07.5)
	≥ 58.5	65 (60.7)		AS	04 (03.7)
FIGO stage	I-II	71 (66.4)	Pelvic LN involvement	Pelvic LN	67 (62.6)
	III-IV	36 (33.6)		PALN	26 (24.3)
				Both	24 (64.5)
Pre-Tx Hb (mg/dL)	Median	11.3 [07.6 – 15.5]	Pre-Tx Hct (%)	Median	33.9 [22.2–45.3]
	<11.45	47 (43.9)		< 34.6	52 (48.6)
	≥11.45	60 (56.1)		≥34.6	55 (51.4)
Pre-Tx WBC count (*)	Median	06.4 [02.4–41.4]	Pre-Tx platelet count (*)	Median	275 [76–451]
	< 6.98	51 (47.7)		< 265.5	51 (47.7)
	≥ 6.98	56 (52.3)		≥ 265.6	56 (52.3)
Pre-Tx NLR	Median	02.3 [00.8–18.6]	Pre-Tx PLR	Median	165.7 [48.7–821.1]
	< 2.33	53 (49.5)		< 136.6	53 (49.5)
	≥ 2.33	54 (50.5)		≥ 136.6	54 (50.5)
Pre-Tx LMR	Median	03.9 [00.6–11.2]	△ NLR	Median	-04.3 [-09.4–10.4]
	< 4.17	53 (49.5)		< 0.0	35 (32.7)
	≥ 4.17	54 (50.5)		≥ 0.0	72 (62.3)
△ PLR	Median	-71.4 [–533.4–715.7]	△ LMR	Median	01.3 [-07.0–03.9]
	< 0.0	21 (19.6)		< 0.0	86 (80.4)
	≥ 0.0	86 (80.4)		≥ 0.0	21 (19.6)

FIGO = International Federation of Gynecology and Obstetrics , Hb = Hemoglobin
AC = Adenocarcinoma, AS = Adenosquamous, SCC = Squamous cell carcinoma.
Hct = Hematocrit, LMR = Lymphocyte-to-monocyte ratio, LN = Lymph node.
*NLR = Neutrophil-to-lymphocyte ratio, PALN = Para-aortic lymph node, * = $x10^3/\mu L$.*
PLR = Platelet-to-lymphocyte ratio, WBC = White blood cell, Pre-Tx = Pre-treatment.

Progression-Free Survival Analysis

Among the 71 patients with stage I and II cancer, 25 patients (35.2%) developed recurrences, whereas 22 patients among the 36 patients with stage III and IV cancer (61.1%) developed recurrences. In the univariate analysis, age, histology, FIGO stage, hematocrit, platelet count, and △NLR had p values < 0.3, which was selected to include the potential, and possibly significant, variables for multivariable analysis. Table 4.2 shows the univariate analysis

for progression-free survival. In the multivariable analysis, age (HR = 2.55, 95% CI 1.38 to 4.70), FIGO stage (HR = 2.49, 95% CI 1.40 to 4.43), and △NLR (HR = 2.91, 95% CI 1.43 to 5.94) were significant predictors of the risk of recurrence. Table 4.3 shows the multivariable analysis for progression-free survival, in which younger age, advanced stage, and a higher △NLR were associated with significantly shorter survival (p = 0.003, p = 0.002, and p = 0.003, respectively).

Overall Survival Analysis

In the univariate analysis, age, FIGO stage, and △NLR had p values < 0.3. Table 4.2 shows the univariate analysis for overall survival. In the multivariable analysis, age (HR = 4.49, 95% CI 1.78 to 11.32), FIGO stage (HR = 3.02, 95% CI 1.32 to 6.90), and △NLR (HR = 3.13, 95% CI 1.18 to 8.27) remained significant predictors of overall survival. Table 4.3 shows the multivariable analysis for overall survival, in which younger age, advanced stage, and a higher △NLR were associated with significantly shorter survival (p = 0.001, p = 0.009, and p = 0.021, respectively).

Table 4.2. Three-years univariate analysis for progression-free survival and overall survival.

		Progression-free survival		Overall survival	
Variables	**Levels**	**Survival probability (95% C I)**	**P-value**	**Survival probability (95% C I)**	**P-value**
Age (years)	< 58.5	40.5 (27.9 - 59.0)	0.076	74.1 (61.7 - 89.0)	0.016
	≥ 58.5	65.1 (53.0 - 79.8)		87.5 (77.6 - 98.7)	
Histology	SCC	54.9 (44.8 - 67.1)	0.110	82.8 (74.2 - 92.4)	0.325
	Others	41.7 (21.3 - 81.4)		87.5 (77.6 - 98.7)	
FIGO stage	I-II	60.8(49.3 - 75.1)	0.007	86.2 (76.9 - 96.6)	0.016
	III-IV	38.8 (25.3 - 59.4)		70.5 (55.9 - 89.0)	
Pelvic LN involvement	Negative	57.1 (42.9 - 76.0)	0.867	83.6 (71.2 - 98.1)	0.016
	Positive	51.3 (39.6 - 66.5)		79.1 (68.5 - 91.3)	
PALN involvement	Negative	55.6 (45.0 - 68.6)	0.507	81.0 (71.6 - 91.6)	0.926
	Positive	46.2 (28.8 - 74.0)		81.0 (65.6 - 99.9)	
Pelvic + PALN involvement	Negative	55.2 (44.7 - 68.1)	0.573	81.2 (71.9 - 91.7)	0.845
	Positive	46.8 (28.9 - 75.5)		80.0 (64.1 - 99.8)	
Hemoglobulin (g/dl)	< 11.45	51.0 (37.7 - 69.1)	0.347	78.3 (65.8 - 93.0)	0.832
	≥ 11.45	55.3 (43.1 - 71.1)		82.9 (72.5 - 94.7)	
Hematocrit (%)	< 34.6	52.1 (39.4 - 68.9)	0.258	75.2 (62.8 - 90.1)	0.558
	≥ 34.6	54.6 (41.8 - 71.5)		85.8 (75.6 - 97.4)	
WBC ($x10^3/\mu L$)	< 06.9	49.3 (36.6 - 66.5)	0.388	83.5 (72.7 - 95.8)	0.892
	≥ 06.9	57.7 (44.8 - 74.1)		78.5 (66.7 - 92.4)	
Platelets ($x10^3/\mu L$)	< 265.5	47.7 (34.8 - 65.5)	0.283	80.0 (68.2 - 93.8)	0.682
	≥ 265.5	58.4 (45.9 - 74.4)		81.7 (70.7 - 94.4)	
NLR	< 02.3	53.9 (41.0 - 70.9)	0.616	81.0 (68.9 - 95.1)	0.727
	≥ 02.3	53.0 (40.3 - 69.7)		80.5 (69.7 - 92.9)	
PLR	< 136.6	52.8 (40.0 - 69.6)	0.827	79.0 (66.6 - 93.6)	0.749
	≥ 136.6	54.2 (41.3 - 71.0)		82.0 (71.3 - 94.2)	
LMR	< 04.1	43.5 (31.1 - 60.8)	0.034	76.9 (65.2 - 90.7)	0.451
	≥ 04.17	62.1 (49.1 - 78.6)		82.4 (70.1 - 96.8)	
△NLR	< 00.0	64.8 (49.7 - 84.5)	0.067	86.4 (74.9 - 99.8)	0.451
	≥ 00.0	47.4 (36.2 - 62.2)		77.7 (66.9 - 90.2)	
△PLR	< 00.0	61.2 (43.4 - 86.4)	0.854	79.9 (64.0 - 99.7)	0.747
	≥ 00.0	51.7 (41.2 - 64.8)		81.1 (71.7 - 91.6)	
△LMR	< 00.0	53.6 (43.2 - 66.5)	0.857	81.7 (72.6 - 92.0)	0.378
	≥ 00.0	52.5 (33.7 - 81.6)		76.5 (58.5 - 100)	

FIGO *= International Federation of Gynecology and Obstetrics ,* ***WBC*** *= White blood cell,* ***SCC*** *= Squamous cell carcinoma.*
LMR *= Lymphocyte-to-monocyte ratio, △****LMR*** *= post treatment LMR - pre treatment LMR,* ***LN*** *= Lymph node.*
NLR *= Neutrophil-to-lymphocyte ratio, △****NLR*** *= post treatment NLR - pre treatment NLR,* ***PALN*** *= Para-aortic lymph node*
PLR *= Platelet-to-lymphocyte ratio, △****PALN*** *= post treatment PLR - pre treatment PLR.*

Table 4.3. Multivariable analysis for progression-free survival and overall survival.

Variables	Levels	HR (95% C I)	β - coefficient	Risk score	P-value
Progression-free survival					
Age	$\geq$ 58.5 vs < 58.5	2.54 (1.38 - 4.69)	0.93	09	0.003
FIGO stage	I-II vs III-IV	2.48 (1.39 - 4.43)	0.91	09	0.002
$\triangle$NLR	$\geq$ 00.0 vs < 00.0	2.91 (1.42 - 5.93)	1.06	10	0.003
Overall survival					
Age	$\geq$ 58.5 vs < 58.5	4.48 (1.77 - 11.32)	1.50	15	0.001
FIGO stage	I-II vs III-IV	3.01 (1.31 - 6.89)	1.10	11	0.009
$\triangle$NLR	$\geq$ 00.0 vs < 00.0	3.13 (1.18 - 8.27)	1.14	11	0.021

Scoring System for Risk Stratification

Based on the prognostic factor analysis for predicting survival, the prognostic risk score for progression-free survival included prognostic factors such as an age < 58.5 years (9 points), FIGO stage III or IV (9 points), and a $\triangle$NLR $\geq$ 0 (10 points) were each assigned scores based on the β-coefficients of the risk factors for survival.

The risk stratification of progression-free survival according to the risk scores was performed as follows: risk score $\leq$ 10, low-risk group (n = 59, 55.1%); risk score 10 < and $\leq$ 19, intermediate group (n=40, 37.4%); risk score > 19, high-risk group (n=8, 7.5%). There were significant differences in the 3-year progression-free survival rates among the low, intermediate, and high-risk patients (70.7%, 40.5%, and 0%, respectively; p < 0.05) (Figure 4.2).

The prognostic risk score for overall survival included prognostic factors such as an age 58.5 years (15 points), FIGO stage III or IV (10 points), and a $\triangle$NLR $\geq$ 0 (10 points); each was assigned scores based on the β-coefficients of the risk factors for survival. The risk stratification of overall survival according to the risk

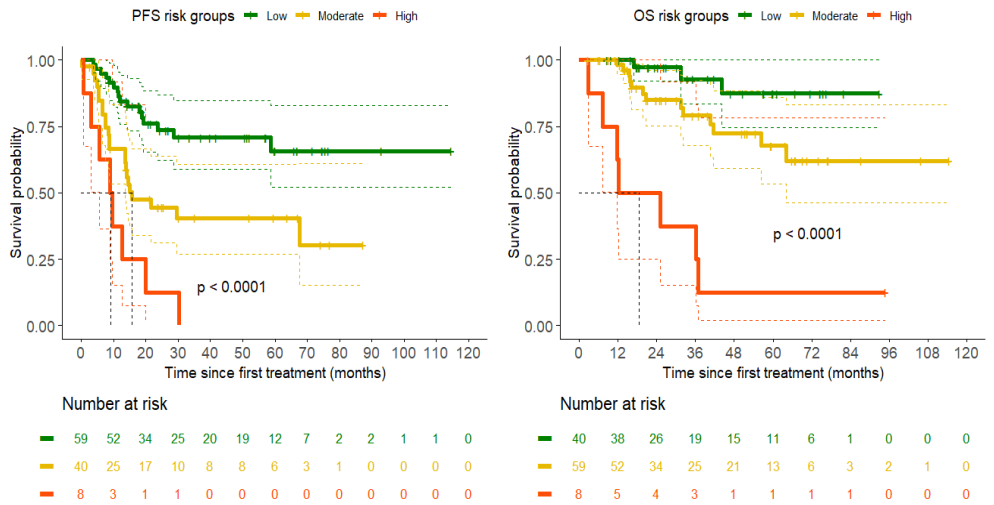

Figure 4.2. *Kaplan–Meier plots of progression-free survival and overall survival according to risk stratification, respectively.*

Scores was performed as follows: risk score $\leq$ 10, low-risk group (n = 40, 37.4%); risk score 10 < and $\leq$ 25, intermediate group (n = 59, 55.1%); risk score > 25, high-risk group (n = 8, 7.5%). There were significant differences in the 3-year overall survival rates among the low, intermediate, and high-risk patients (92.8%, 79.1%, and 37.5%, respectively; $p < 0.05$) (Figure 4.2). Table 4.3 shows the risk-scoring algorithm based on the β-coefficients of the risk factors obtained from Cox proportional hazards model for survival. The concordance indices for progression-free survival and overall survival were 0.663 and 0.736, respectively. Figure 4.3 shows the calibration curves for 3-year survival.

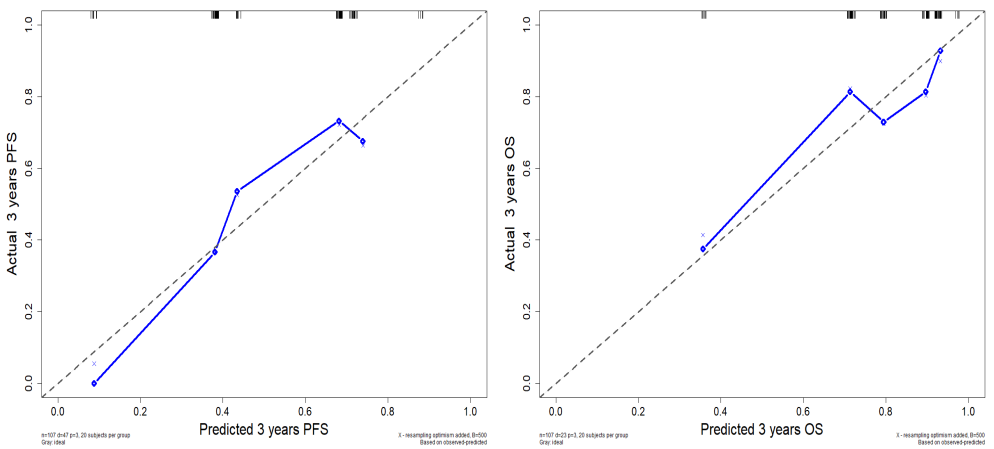

Figure 4.3. *Calibration curve of progression-free survival and overall survival, respectively. The plot illustrates how the predicted probabilities from the model compare with the actual patient outcomes. Gray line represents the performance of an ideal prediction model, and blue the presented prediction models.*

Discussion

This study retrospectively analyzed prognostic factors predicting survival in cervical cancer patients treated with chemoradiation. As a result, we found that the prognostic significant factors for survival were age, FIGO stage, and △NLR. Based on these results, risk scores for survival were generated from the β-coefficients of the risk factors (12). Furthermore, risk stratification for survival according to the risk scores were performed to identify the high-risk group of cervical cancer patients treated with chemoradiation. In cervical cancer, the age and FIGO stage at diagnosis have been reported as significant prognostic factors for predicting survival (15–17). Recently, there has been a report that the hematologic parameters, such as a low lymphocyte percentage and a high NLR, were associated with younger age, an advanced stage, and other parameters (17). Our findings are similar to the results of the Onal et al. (18) study , which showed that younger age, a higher FIGO stage, and an increased pre-treatment NLR were associated with poor survival. However, in our study, △NLR rather than pre-treatment NLR, showed a prognostic significance for survival. Previous studies showed that systemic inflammatory markers such as NLR, PLR, and LMR are prognostic factors for various types of carcinomas, including cervical cancer (6–8, 18, 19). The results of the studies analyzing the correlations between hematologic parameters and the prognosis of patients who underwent treatments are summarized in Table S4.1 of the supplementary material. In contrast to these results, in our study, the NLR, PLR, and LMR did not show significance in predicting survival, but the △NLR showed statistical significance, demonstrating the value of the combination both of the pre- and post-treatment NLR. Because of the different treatment modalities used and the retrospective nature of the study, direct comparisons and interpretations of different studies are somewhat difficult. Furthermore, in most studies, the hematologic values were assessed during the pre-treatment period, and △NLR was rarely analyzed.

For the association between the hematologic parameters and the clinical outcome of cancer patients, several explanations have been postulated. Neutrophils produce cytokines and vascular endothelial growth factor, which play an important role in angiogenesis(20). Conversely, circulating lymphocytes participate in preventing the proliferation and metastasis of tumor cells by producing cytokines (21). Therefore, when taken together, these results may indicate that the NLR may reflect the immune and inflammatory responses. Additionally, patients with cervical cancer with an increased pre-treatment NLR may experience relative neutrophilic leukocytosis and lymphocytopenia, which suggests the presence of a pro-tumor inflammatory response that is likely to result in poor survival. Similarly, a change in the post-treatment NLR might reflect changes in the immune and inflammatory responses (22). If the NLR increases after chemoradiation, this indicates the state of the pro-tumor inflammatory response.

Through the assessment of each independent risk factor associated with survival, we established a prediction scoring system for cervical cancer patients treated with chemoradiation. The distinction between the different risk stratification based on this scoring system might provide further guidance for individual treatment and optimized treatment plans. Furthermore, considering the number of events in progression-free survival and overall survival, we believe the results of this study are not overfitted and are reproducible according to the "one in ten rule", which states that one predictive variable can be analyzed for every ten events. This rule is a criterion for how many predictor variables can be estimated from the data (23).

Therefore, we have also tried to perform the multivariable analysis for survival based on all the independent variables including pre-treatment NLR, PLR, and LMR, although these parameters were were not shown as significant parameters or p values < 0.3 in univariate analysis, but the △NLR only showed statistical significance among the hematologic parameters. Based on the results of this study, we hypothesized that the △NLR is the strongest predictor of survival among the other hematologic variables such as the pre-treatment NLR or post-treatment NLR.

In this study, for the assessment of treatment response, including lymph node involvement and the detection of distant metastatic disease, some patients underwent PET/CT, which has been reported as resulting in a high false-positive rate, especially for the detection of lymph node metastasis (24). Although lymph node involvement was not a prognostic significant parameter, further analysis based on patient staging with the 2018 FIGO staging system is needed.There were some limitations to this study. First, we analyzed patients who were retrospectively enrolled and there was a relatively small sample size. Second, this research was conducted based on the patients at a single institution. Besides, we could not perform the validation analysis based on another cohort due to the above reasons. We believe that further study with a well-designed clinical trial with more patients enrolled at multicenter institutions could confirm the exact significance of the results of the study.

Conclusion

In conclusion, for survival, as a hematologic parameter, the △NLR was an independent prognostic factor. In addition, as clinical parameters, age and FIGO stage were independent prognostic factors for cervical cancer patients treated with chemoradiation . Based on these results, we developed a risk stratification method based on a risk scoring system for cervical cancer patients who received chemoradiation.

Supplemental materials

Table 4.1. Comparison of the studies that analyzed the correlation between hematologic parameters and the prognosis of cervical cancer patients who underwent treatment.

				Hazard Ratio (95% Confidence Interval)	
Study	**Patients (n)**	**Treatment**	**Variable**	**Overall survival**	**Progression-free survival**
Nakamura et al. (6)	32	CCRT	Pre-treatment PLR	4.20 (1.15 - 15.26)	-
Mizunuma et al. (7)	56	RT or CCRT	Pre-treatment NLR	2.80 (0.83 - 9.34)	1.53 (1.19 - 1.97)
Chen et al. (8)	485	Surgery	DSI	-	1.39 (1.01 - 1.93)
			FIGO stage	-	1.32 (0.97 - 1.81)
			LN metastasis	1.59 (1.29 - 1.97	1.73 (1.457 - 2.06)
			Pre-treatment LMR	0.41 (0.244 - 0.71)	0.43 (0.27 - 0.69)
Onal et al. (18)	235	CCRT	Age	1.01 (1.00 - 1.03)	-
			LN metastasis	2.62 (1.70 - 4.02)	2.98 (1.91 - 4.37)
			Pre-treatment NLR	3.32 (1.90 - 5.79)	3.57 (2.10 - 6.08)
Lee et al. (19)	1061	Surgery (RH) or RT or CCRT	Age	1.02 (1.00 to 1.04)	-
			Pre-treatment NLR	1.19 (1.13 to 1.25)	1.13 (1.08 to 1.18)
Present study	107	CCRT	Age	4.48 (1.77 - 11.32)	2.54 (1.38 - 4.69)
			FIGO stage	3.01 (1.31 - 6.89)	2.48 (1.39 - 4.43)
			△NLR	3.13 (1.18 - 8.27)	2.91 (1.42 - 5.93)

FIGO = International Federation of Gynecology and Obstetrics , CCRT = Concurrent Chemoradiation Therapy, LMR = Lymphocyte-to-monocyte ratio, DSI = Depth of Stromal Invasion, LN = Lymph Node. NLR = Neutrophil-to-lymphocyte ratio, △NLR = Post treatment NLR - Pre treatment NLR, RH = Radical Hysterectomy. PLR = Platelet-to-Lymphocyte Ratio, RT = Tadiation Therapy.

Bibliography

[1] Sun-Seog Kweon. Updates on cancer epidemiology in korea, 2018. *Chonnam medical journal*, 54(2):90, 2018.

[2] Hyoung Jin Kim and Hong-Jin Kim. Current status and future prospects for human papillomavirus vaccines. *Archives of pharmacal research*, 40(9):1050–1063, 2017.

[3] Y Cho, KH Kim, HI Yoon, GE Kim, and Yong Bae Kim. Tumor-related leukocytosis is associated with poor radiation response and clinical outcome in uterine cervical cancer patients. *Annals of oncology*, 27(11):2067–2074, 2016.

[4] Myong Cheol Lim, Young-Joo Won, Mun Jung Ko, Miseon Kim, Seung-Hyuk Shim, Dong Hoon Suh, and Jae-Weon Kim. Incidence of cervical, endometrial, and ovarian cancer in korea during 1999–2015. *Journal of gynecologic oncology*, 30(1), 2019.

[5] IL Atahan, C Onal, Enis Ozyar, F Yiliz, U Selek, and F Kose. Long-term outcome and prognostic factors in patients with cervical carcinoma: a retrospective study. *International Journal of Gynecologic Cancer*, 17(4), 2007.

[6] Keiichiro Nakamura, Takeshi Nishida, Tomoko Haruma, Junko Haraga, Chiaki Omichi, Chikako Ogawa, Tomoyuki Kusumoto, Noriko Seki, Hisashi Masuyama, and Yuji Hiramatsu. Pretreatment platelet-lymphocyte ratio is an independent predictor of cervical cancer recurrence following concurrent chemoradiation therapy. *Molecular and clinical oncology*, 3(5):1001–1006, 2015.

[7] Makito Mizunuma, Yoshihito Yokoyama, Masayuki Futagami, Masahiko Aoki, Yoshihiro Takai, and Hideki Mizunuma. The pretreatment neutrophil-to-lymphocyte ratio predicts therapeutic response to radiation therapy and concurrent chemoradiation therapy in uterine cervical cancer. *International journal of clinical oncology*, 20(5):989–996, 2015.

[8] Liang Chen, Fang Zhang, Xiu-gui Sheng, and Shi-qian Zhang. Decreased pretreatment lymphocyte/monocyte ratio is associated with poor prognosis in stage ib1–iia cervical cancer patients who undergo radical surgery. *OncoTargets and therapy*, 8:1355, 2015.

[9] Yang Chen, Huan Yan, YanRong Wang, Yan Shi, and GuangHai Dai. Significance of baseline and change in neutrophil-to-lymphocyte ratio in predicting prognosis: a retrospective analysis in advanced pancreatic ductal adenocarcinoma. *Scientific reports*, 7(1):1–9, 2017.

[10] Olivier Trédan, Isabelle Ray-Coquard, Gisèle Chvetzoff, Paul Rebattu, Agathe Bajard, Sylvie Chabaud, David Pérol, Chadi Saba, Florent Quiblier, Jean-Yves Blay, et al. Validation of prognostic scores for survival in cancer patients beyond first-line therapy. *BMC cancer*, 11(1):1–9, 2011.

[11] Wei Peng, Chuan Li, Tian-Fu Wen, Lv-Nan Yan, Bo Li, Wen-Tao Wang, Jia-Yin Yang, and Ming-Qing Xu. Neutrophil to lymphocyte ratio changes predict small hepatocellular carcinoma survival. *journal of surgical research*, 192(2):402–408, 2014.

[12] Hemalkumar B Mehta, Vinay Mehta, Cynthia J Girman, Deepak Adhikari, and Michael L Johnson. Regression coefficient–based scoring system should be used to assign weights to the risk index. *Journal of clinical epidemiology*, 79:22–28, 2016.

[13] IBM Corp. Ibm spss statistics for windows. URL https://hadoop.apache.org.

[14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[15] C Rivard, E Stockwell, J Yuan, R Isaksson Vogel, and MA Geller. Age as a prognostic factor in cervical cancer: A 10-year review of patients treated at a single institution. *Gynecologic Oncology*, 141:102, 2016.

[16] Jian Yang, Haoyang Cai, Zhi-Xiong Xiao, Hangyu Wang, and Ping Yang. Effect of radiotherapy on the survival of cervical cancer patients: an analysis based on seer database. *Medicine*, 98(30), 2019.

[17] Myung-Hwa Jeong, Hyungjoo Kim, Tae-Hun Kim, Moon-Hong Kim, Beob-Jong Kim, and Sang-Young Ryu. Prognostic significance of pretreatment lymphocyte percentage and age at diagnosis in patients with locally advanced cervical cancer treated with definite radiotherapy. *Obstetrics & gynecology science*, 62(1):35, 2019.

[18] Cem Onal, Ozan C Guler, and Berna A Yildirim. Prognostic use of pretreatment hematologic parameters in patients receiving definitive chemoradiotherapy for cervical cancer. *International Journal of Gynecologic Cancer*, 26(6), 2016.

[19] Yoo-Young Lee, Chel Hun Choi, Ha-Jeong Kim, Tae-Joong Kim, Jeong-Won Lee, Je-Ho Lee, Duk-Soo Bae, and Byoung-Gie Kim. Pretreatment neutrophil: lymphocyte ratio as a prognostic factor in cervical carcinoma. *Anticancer research*, 32(4):1555–1561, 2012.

[20] Yoka H Kusumanto, Wendy A Dam, Geke AP Hospers, Coby Meijer, and Nanno H Mulder. Platelets and granulocytes, in particular the neutrophils, form important compartments for circulating vascular endothelial growth factor. *Angiogenesis*, 6(4): 283–287, 2003.

[21] Pei-Rong Ding, Xin An, Rong-Xin Zhang, Yu-Jing Fang, Li-Ren Li, Gong Chen, Xiao-Jun Wu, Zhen-Hai Lu, Jun-Zhong Lin, Ling-Heng Kong, et al. Elevated preoperative neutrophil to lymphocyte ratio predicts risk of recurrence following curative resection for stage iia colon cancer. *International journal of colorectal disease*, 25(12):1427–1433, 2010.

[22] Youngjoo Lee, Sun Hye Kim, Ji-Youn Han, Heung Tae Kim, Tak Yun, and Jin Soo Lee. Early neutrophil-to-lymphocyte ratio reduction as a surrogate marker of prognosis in never smokers with advanced lung adenocarcinoma receiving gefitinib or standard chemotherapy as first-line therapy. *Journal of cancer research and clinical oncology*, 138 (12):2009–2016, 2012.

[23] Peter Peduzzi, John Concato, Elizabeth Kemper, Theodore R Holford, and Alvan R Feinstein. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology*, 49(12):1373–1379, 1996.

[24] Laurie L Brunette, Shahram Bonyadlou, Lingyun Ji, Susan Groshen, Dena Shuster, Arjun Mehta, Richard Sposto, Koji Matsuo, Yvonne G Lin, and Lynda D Roman. Predictive value of fdg pet/ct to detect lymph node metastases in cervical cancer. *Clinical nuclear medicine*, 43(11):793, 2018.

CHAPTER 5

OVERALL SURVIVAL NOMOGRAM FOR PATIENTS WITH SPINAL BONE METASTASES

Authors:
***Biche Osong**, Ilknur Sanli, Paul C. Willems, Leonard Wee, Andre Dekker, Seok Ho Lee, Johan van Soest*

***Adapted from**: Clinical and Translational Radiation Oncology (ctRO)*
***DOI**: 10.1016/j.ctro.2021.02.010*

Abstract

Purpose
This study aimed to develop a nomogram for predicting 1, 3, and 6-months overall survival of patients with metastatic spinal tumors and provide a tool for easy use and integration into the clinic.

Summary of background
About 30 - 70% of patients with a primary tumor have a metastatic spinal disease at autopsy. Spinal metastasis forms a significant disease burden, with the spine being the third most common site for metastases.

Patients and Methods
We reviewed a total of 250 patients with spinal bone metastases admitted to our institution between January 2014 to April 2016 for this study. Only breast, prostate, colon, rectal, and lung cancer were included as primary tumor sites, reducing the number of patients for this study to 195. A 5-fold cross-validation Cox proportional hazard regression model with the lasso penalty was employed for the feature selection process before establishing the prognostic nomogram. The discrimination developed nomogram was measured by the concordance index (C-index). A bootstrap calibration plot was used to ascertain the model's accuracy.

Results
Six independent prognostic factors, including age, the presence of visceral metastasis, spinal cord compression, brain metastasis, WHO performance status, and primary tumor, were identified during the feature selection process for building the nomogram with the addition of gender. The calibration curves for 1, 3, and 6-month overall survival showed good agreement between the predictive and the actual probabilities. The nomograms mean C-index was 0.720 (95% CI, 0.683 – 0.757)

Conclusion
We established a user-friendly nomogram to predict survival probability for patients with spinal bone metastasis. We provided a tool for flexible and easy usage, which can help physicians with their decision-making process and the individualized care planning of SBM patients.

Introduction

Tumor metastasis is the leading cause of morbidity and mortality in cancer patients (1, 2). The spine is the third most common site for cancer cells to metastasize after lung and liver, and 30 - 70% of patients with a tumor have metastatic spinal disease at autopsy (1, 3–5).

Primary tumors of the breast, prostate, thyroid, lung, gastrointestinal (GI), and kidney are the most common to metastasize to the spine (1, 3–5). Within the spinal column, metastases are more commonly found in the thoracic spine, followed by the lumbar spine, while the cervical spine is the least likely location to find metastasis.

Spinal bone metastases (SBM) account for over 70% of all osseous metastases and are slightly more common in men than in women. Adults between the ages of 40 and 65 are affected more than any other age group (4–6). The prognosis of SBM is abysmal and heavily depends on the primary tumor (7). Only 10 to 20 percent of the diagnosed patients have survival of more than two years, which implies that caregivers should tailor treatment based on an individual patient profile for an optimal outcome.

Graphical tools such as nomograms that can be used to estimate an event's probability by assigning scores to each important risk factor known to impact the events of interest combined with a prediction model can be used in such a situation. Since nomograms can estimate patient-specific probability of an outcome, they are an excellent decision support system for clinicians and caregivers.

Numerous nomograms have been developed for different cancer-specific outcomes (8–13). and thanks to the technological advancements in the oncological field in the last decade, some of these nomograms have been digitalized (14). However, until now, no prognostic nomogram has been established for SBM.

Therefore, this study aims to develop a nomogram with a user-friendly digital interface that can estimate the 1, 3, and 6-month overall probabilities of survival for patients with SBM and guide individualized patient management decisions.

Materials and methods

Between January 2014 and April 2016, we retrospectively collected a series of 250 cancer patients treated for SBM from the electronic medical record (EMR) system at Maastro Clinic, Maastricht, The Netherlands, after acquiring approval from the internal review board. All the patients received radiotherapy for their metastatic tumor. We extracted the following patient demographics and clinical information age, sex, WHO performance status, pathological fracture, spinal cord compression, number of spinal metastases, extra spinal metastases, visceral metastases, brain metastases, lymphatic metastases, pain score, and primary tumor for this analysis. We only included patients with a primary tumor of the breast, prostate, colon, rectum, or lung in this study. Overall survival (OS) at 1, 3, and 6 months was defined as the primary outcome of interest. The OS was calculated as the time difference between the date of diagnosis and the date of death or last follow-up.

LASSO

The least absolute shrinkage and selection operator (LASSO) method which performs the **L1** regularization (L1 norm) was used to select the important variables to build the nomogram. The regularization parameter (λ) imposes a penalty on the absolute value of the magnitude of the coefficients, leading to some variables shrinking to zero depending on the λ value. The λ parameter has two optimal values called minimum criteria (λ_{min}) and 1 standard error of the minimum criteria (λ_{1-SE}) which are the smallest and largest λ values for each cross-validation run, respectively. The λ_{min} value gives a model which has the minimum mean cross-validated error while λ_{1-SE} gives a model which is 1 standard error away from the minimum error across all the folds(15).The λ parameter have two optimal values called minimum criteria (λ_{min}) and 1 standard error of the minimum criteria (λ_{1-SE}) which are the smallest and biggest λ values respectively

Statistics

Descriptive statistics and data visualization were applied to understand and detect the data sets underlying patterns such as missing information and possible outlying values. A 5-fold cross-validation Cox proportional hazard regression model with the least absolute shrinkage and selection operator (LASSO) penalty was used to select features that can predict survival for patients with SBM. The optimal λ values which compromise model complexity and performance were determined using the **cv.glmnet** function. Variables with a non-zero coefficient under the λ_{min} value were used to fit a multivariate Cox proportional hazard regression model. The fitted multivariate Cox proportional hazard regression model was translated to a nomogram for visualization using the **nomogram** function from the rms package (16) The accuracy of the nomogram on a repeated (R = 10) 5-fold cross-validation was measured based on the concordance index (C-index) value with a C-index of 1 indicating a perfect nomogram and a C-index of 0.5 implying the nomogram is

as reliable as tossing a coin. An internal bootstrap (B = 500) correction plot of observed against nomogram-predicted survival probability was used to calibrate the nomogram at the different time points of interest.

The linear predictors (LP) which are the linear combination of the coefficients of the variables in the nomogram were discretized to create the survival risk groups. Survival difference was visualized and tested using Kaplan-Meier plots and log-rank test, respectively. To evaluate the models' ability to classify future patients into the different risk groups, we compared the predicted mean survival curves for each of the risk groups with the true Kaplan-Meier survival curves of each risk group by overlaying the two plots. All statistical analyses were performed using **R** software (17) and the glmnet package (18) was used for variable selection and model fitting process.

Results

A total of 250 patients with SBM were identified at Maastro Clinic. Of these patients, 195 had a primary tumor of the breast, prostate, colon, rectum, or lung. One patient with missing WHO performance status was excluded from this analysis. The variable 'pain score' was excluded from the study due to its high percentage (45%) of missing information. The median age of patients in this study was 69 (39 - 92) years. There was no statistical survival difference between surviving and non-surviving patients for all considered variables but visceral metastasis and the primary tumor. The median follow-up time for this study was 46.78 (37.03 - 56.34) months with a 1, 3, and 6-month overall survival probability of 88%, 67%, and 53%, respectively. Table 5.1 shows the general patient characteristics for this study.

Table 5.1. General characteristics for surviving and non surviving patients

Variable	Levels	Survivors	Non-Survivors	p-value
Age at RT in years	Mean (sd)	67.8 (8.8)	68.9 (10.4)	0.651
Sex	Female	10 (66.67%)	80 (44.40%)	0.097
	Male	05 (33.33%)	100 (55.60%)	
WHO performance score	Active	01 (6.67%)	05 (2.78%)	0.854
	Restricted	07 (46.67%)	69 (38.33%)	
	Self-care	05 (33.33%)	74 (41.11%)	
	Bed-bound	02 (13.33%)	31 (17.22%)	
	Missing	00 (0.00%)	01 (0.56%)	
Pathological fracture	Yes	11 (73.33%)	141 (78.33%)	0.654
	No	04 (26.67%)	39 (21.67%)	
Spinal compression	No	14 (93.33%)	142 (78.89%)	0.179
	Yes	01 (6.67%)	38 (21.11%)	
Number spinal metastases	One	03 (20.00%)	33 (18.33%)	0.981
	Two	03 (20.00%)	39 (21.67%)	
	Three +	09 (60.00%)	108 (60.00%)	
Extra spinal bone metastases	No	04 (26.67%)	41 (22.78%)	0.731
	Yes	11 (73.33%)	139 (77.22%)	
Visceral metastases	Absent	13 (86.67%)	109 (60.56%)	0.045
	Present	02 (13.33%)	71 (39.44%)	
Brain metastases	Absent	00 (0.00%)	10 (5.56%)	0.348
	Present	15 (100%)	170 (94.44%)	
Lymphatic metastases	Absent	09 (60.00%)	102 (56.67%)	0.802
	Present	06 (40.00%)	78 (43.33%)	
Pain score	No pain	00 (0.00%)	05 (2.78%)	0.431
	Mild	01 (6.67%)	06 (3.33%)	
	Moderate	05 (33.33%)	35 (19.44%)	
	Severe	07 (46.67%)	76 (42.22%)	
	Missing	02 (13.33%)	58 (32.22%)	
Primary tumor	Breast	10 (66.67%)	35 (19.44%)	< 0.05
	Prostate	05 (33.33%)	50 (27.78%)	
	Lung	00 (0.00%)	70 (38.89%)	
	Colon	00 (0.00%)	14 (07.78%)	
	Rectum	00 (0.00%)	11 (06.11%)	

WHO = World Health Organization , sd= standard deviation

Figure 5.1A shows a plot of the model performance (C-index) against the log values of the different λ used in the cross-validation process for variable selection. The values at the top of the plot indicate the number of non-zero variables in the model for a particular λ value and the performance of the said model can be read on the y-axis. Based on the selected λ_{min} value from the repeated 5-fold cross-validation of the LASSO Cox proportional hazard regression model, the 11 considered variables were reduced to 6 potential predictors (age,

spinal cord compression, brain metastasis, visceral metastasis, WHO performance status, and primary tumor) with a non-zero coefficient. Figure 5.1B shows the coefficients of the 11 variables represented by different colors against the log(λ) values. The vertical dotted gray line was drawn at the selected λ_{min} value which resulted in the 6 variables with nonzero coefficients.

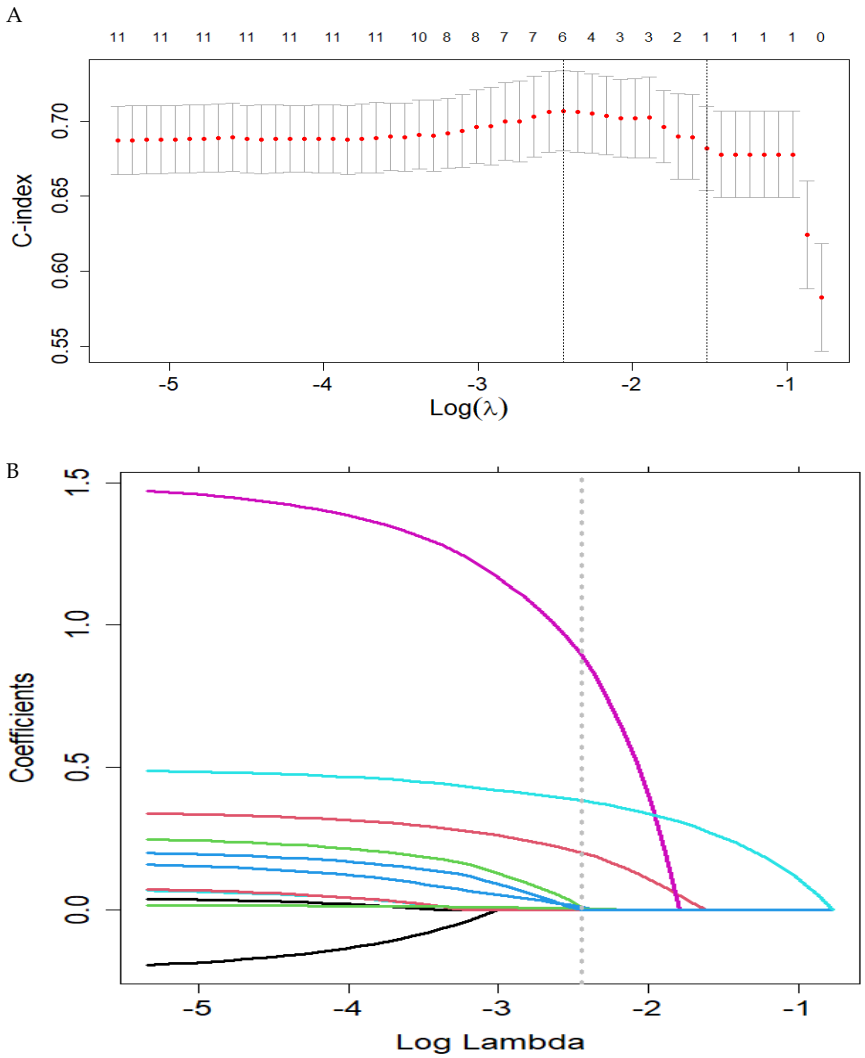

Figure 5.1. *Variable selection using the LASSO Cox proportional hazard regression model. [A] Selection plot of the tuning parameter (λ) for the LASSO model on the repeated 5-fold cross-validation. The C-index values were plotted against the log(λ) values. Dotted vertical lines are drawn at the optimal λ values λ_{min} and λ_{1-SE} respectively. [B] Profile plot of the LASSO coefficient against the log(λ) sequence for the 11 considered variables. The dotted gray line represents the selected λ_{min} value (0.0895) which gives a log (λ_{min}) of -2.413.*

The fitted multivariate Cox proportional hazard regression model with the selected variables was translated to the prognostic nomogram shown in Figure 5.2. The variable sex was included in the model thou not selected based on the chosen λ value because it is known to

be an important factor based on literature. Also, The Kaplan-Meier plot for sex (Figure S5.2 supplemental material) showed a significant survival difference. The mean C-index and the 95% confidence interval(CI) of the nomogram was 0.720 (0.683 – 0.757).

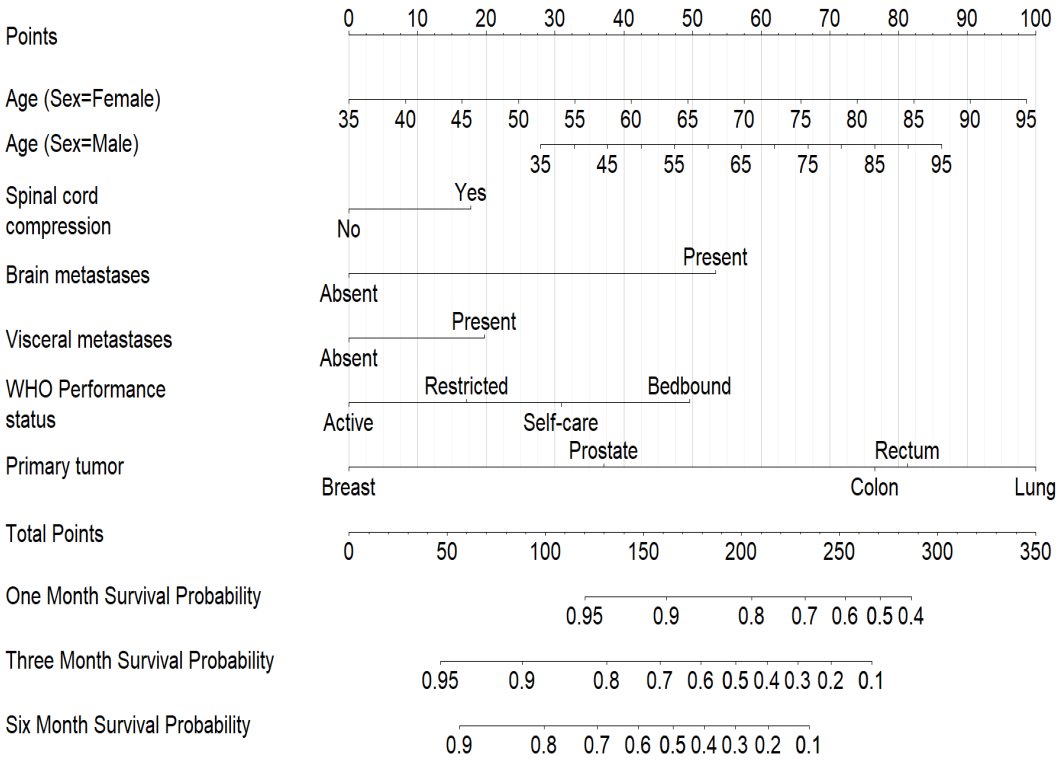

Figure 5.2. *Developed nomogram to predict 1, 3, and 6-months overall survival for metastatic spinal bone patients using seven clinical characteristics. To use the nomogram, locate the patient's variable on the corresponding axis, draw a vertical line to the points axis, sum the points, and draw a vertical line from the total points axis to the 1, 3, or 6 - months overall survival probability axis.*

We have also provided a user-friendly online version of this nomogram to facilitate its widespread use by physicians and researchers (Link[1]). The Web application allows predicted survival probabilities and curves for each input information to be stacked making comparison easier.

To evaluate the developed nomogram, we presented its performance in predicting 1, 3, and 6 months overall in terms of discrimination by plotting the actual survival probabilities against the nomogram-predicted probabilities. This plot shows the similarity between the predicted probabilities and the observed probabilities, with all points falling precisely on the perfect model's diagonal line. The calibration curve in figure 5.3 reveals good agreement between the predictions of the nomogram and observation.

[1]https://bich.shinyapps.io/SpinalMets/

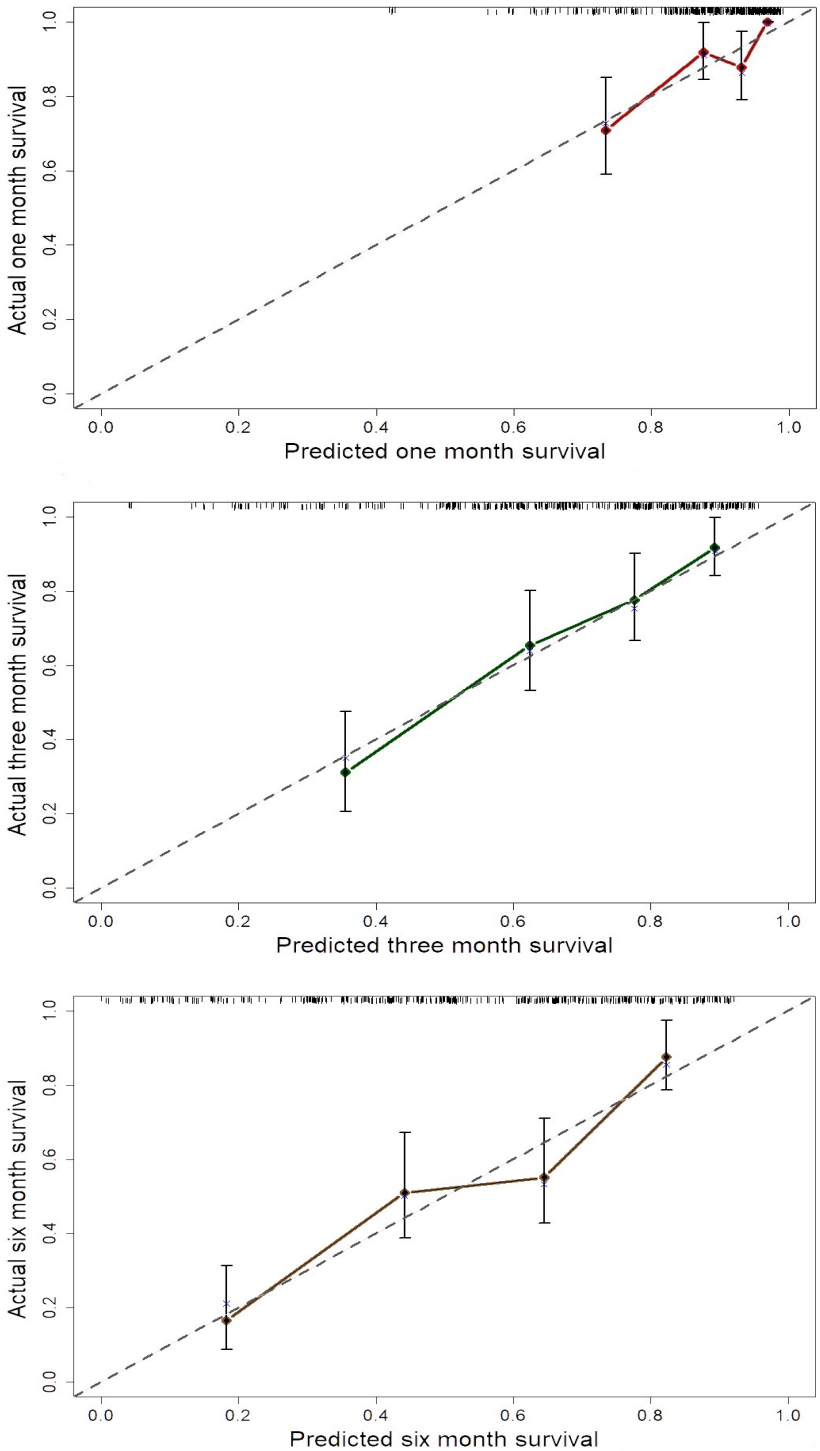

Figure 5.3. *SBM overall survival nomogram calibration plots for 1, 3, and 6 months, respectively. The nomogram-predicted overall survival is plotted on the x-axis, and the actual overall survival is plotted on the y-axis. The dashed line represents the ideal fit where the nomogram-predicted probability matches the observed probability. The vertical solid lines represent the 95% confidence interval.*

The nomograms' ability to discriminate between patients based on their survival probability was evaluated by first making a histogram of the linear predictors, as shown in figure 5.4 with higher values indicating poor prognosis. The linear predictors were then discretized into three risk groups with cutoff values at the 25^{th} and 75^{th} percentile, as shown on the plot. We considered patients between the cutoff values to have a moderate risk of death. Patients below and above the 25^{th} and 75^{th} percentile values were considered to have a lower and higher risk of death, respectively.

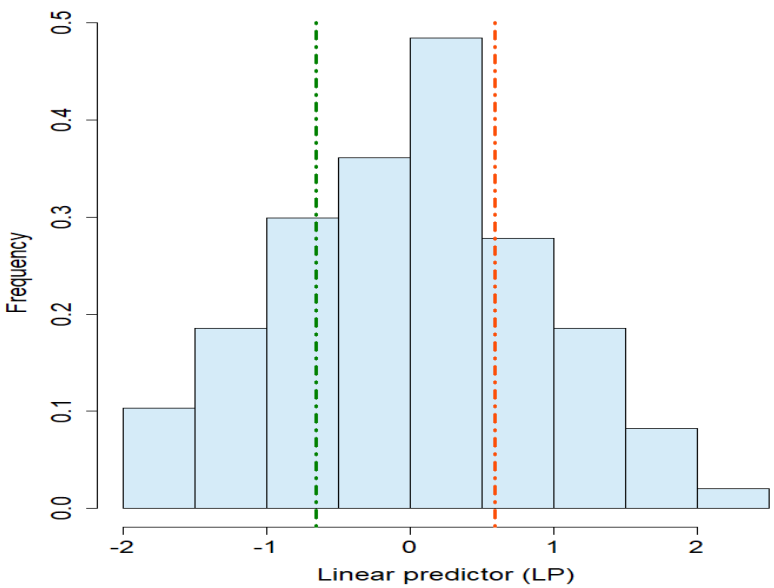

Figure 5.4. *Histogram of the linear predictor extracted from the nomogram. The vertical lines indicates the 25^{th} (green), and 75^{th} (red) percentile respectively.*

The percentages of patients in the three risk groups are 25.3%, 49.4%, and 25.3%, respectively. The Kaplan-Meier curves for overall survival stratified by the risk groups, as shown in figure 5.5, agree with the c-index value and calibration plots, indicating that the nomogram has some discriminating power as the three curves are significantly separated with a p-value $<$ 0.005. Patients in the high-risk group had a median survival time of 1.77 (0.92 - 3.98) months and the moderate group had 6.90 (2.66 - 15.21) while the low-risk group had 25.72 (13.40 - 45.47) months as shown in figure 5.5.

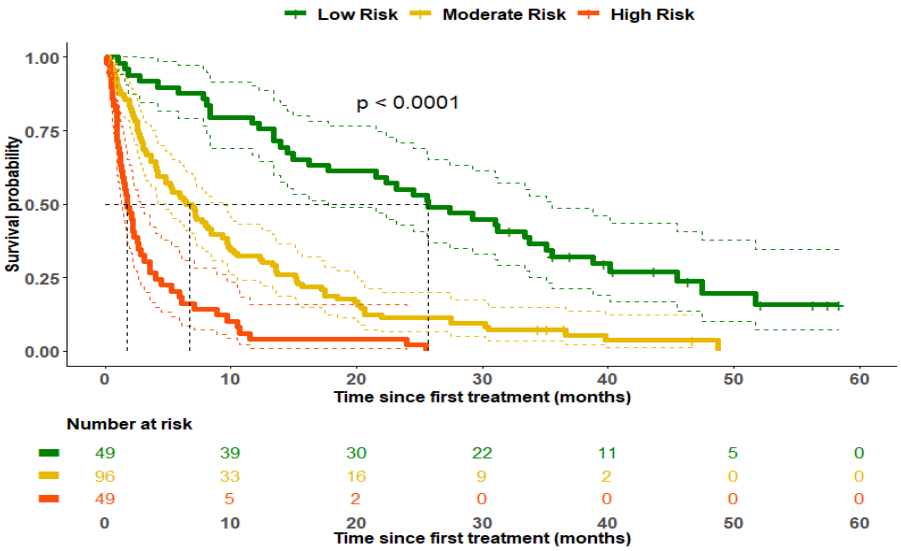

Figure 5.5. *The Kaplan-Meier survival curve for the low, moderate, and high-risk groups based on the percentile cutoff values.*

To further evaluate the nomogram's performance, we compared the predicted mean survival curves for each of the risk strata with the Kaplan-Meier survival curves, as shown in figure 5.6.

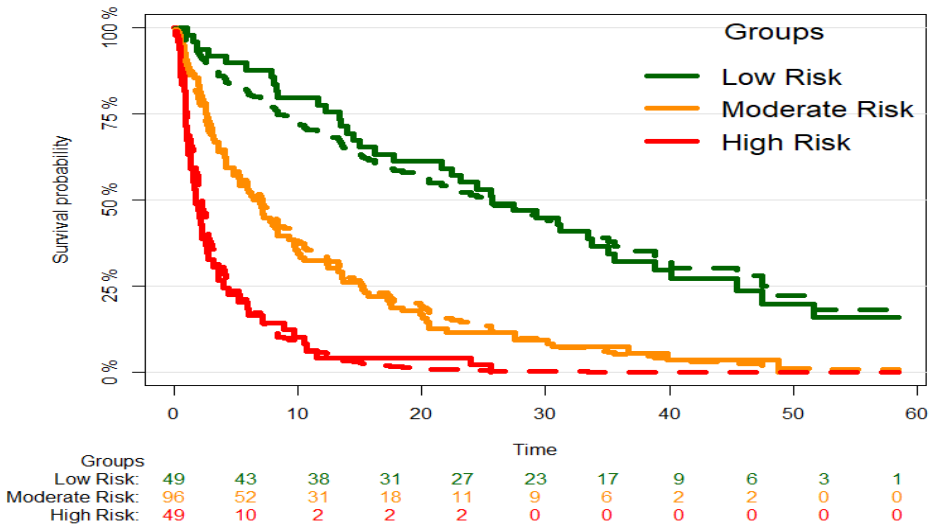

Figure 5.6. *Comparison of predicted mean survival curves (dotted lines) and stratified Kaplan-Meier (solid lines) for the different risk groups.*

Figure 5.6 indicates that the nomogram is well-calibrated given the close similarity between the predicted (dotted lines) and actual (solid lines) survival curve for all except the low-risk group, where the model slightly under-predicts at the beginning and over-predict over time.

Discussion

The disease burden and mortality rate of SBM have opened up intriguing research possibilities in the field, focusing on improving patients' quality of life via a personalized treatment procedure for an optimal outcome. Despite the significant progress in understanding tumor metastasis and the underlying mechanisms, the precise process remains complicated with multiple sequential and interrelated biochemical events, which still need elucidation.

The treatment choice for spinal metastases depends on correctly localizing the affected vertebra(e), the patient's priorities for treatment, and other individual patient characteristics. However, no therapy has proven to increase the life expectancy of these patients (5). Hence, treatment aims to improve quality of life, spinal cord compression, relieve pain or prevent a vertebral collapse (19). Therefore, assessing a patient's prognosis before treatment is very pivotal for an optimal treatment selection. That is, caregivers should tailor treatment based on each patient's desires and overall prognosis.

Renowned prognostic scoring systems (Bauer, Tokuhashi, Tomita, van der Linden, Sioutos, Katagiri, and NESMS) have been developed to assist clinicians and care providers in determining the survival prognosis of metastatic spine tumor patients for an optimal therapeutic choice (20–28). In contrast to this study, none of these scoring systems include demographic features such as age and sex. Logically, these variables should be included in any scoring system given that these variables have been proven from literature to be associated with SBM survival, as the disease is more common in men than women as well as in elderly patients as compared to the younger population (4–6, 29). More to this, men are more at risk of developing a spinal disease than women since men are more susceptible to developing a primary tumor than women (30, 31)

Yang et al. (32), Liu et al. (33) and Pereira et al. (34) have previously developed nomograms to support the personalized predictions of survival probability for patients with spinal metastasis disease from non-small cell lung cancer (NSCLC), colorectal cancer, and operable patients respectively. These nomograms did consider age, sex, performance status, primary tumor, visceral, and brain metastasis as significant prognostic factors associated with spine metastasis survival, which are in concordance with this study. However, none of these studies have considered including both age and sex in the same nomogram. This assumes all patients have an equal risk of dying from the disease irrespective of their age, sex, or both variables despite the sea of literature supporting these differences (4–6, 29, 31, 35, 36) especially when more than one primary tumor is considered (Figure S5.1 supplemental material). This variable omission implies the predicted survival probabilities from such nomograms are less personalized.

We developed a nomogram with seven variables, including an interaction between age and sex, to improve previously developed scoring systems. The developed nomogram captures the age effect within the sex variable as there is over 15 points survival difference between males and females of the same age. From the nomogram, women have relatively better survival than men before 75 years. However, after 75 years, the reverse is seen with men having a somewhat better survival than women. The proposed nomograms have relatively good c-indexes of 0.72 (95% CI, 0.683 – 0.757) and perform well in calibration. A digital version of the nomogram is also provided for easy insertion into the treatment workflow for better decision-making in managing spinal metastases and offering practical guidance to caregivers.

All the existing scoring systems for SBM known to us are between 1 - 24 months. The digital version of the present nomogram can make predictions at any given time point as low as half a month. Besides the survival probability, it also provides the confidence interval of the predicted survival probability and a personalized survival curve, which gives the caregiver more insights to determine the optimal therapeutic strategy for a patient, such as, e.g., stereotactic body radiation therapy (SBRT). The personalized survival curve could serve as a good starting point for shared decision-making between patients and caregivers.

The present nomogram might be a suitable tool for clinical assistance; however, the performance is still not optimal due to some limitations. The nomogram's clinical reliability could not be evaluated at the moment, given the study's single-center nature. However, we performed a thorough internal validation (bootstrap) and planned to do a proper external validation to ascertain the nomogram's clinical usefulness. A direct comparison between our developed nomogram and the other nomograms was not possible due to population differences. However, Liu et al. (33) and Pereira et al. (34) did consider hematological parameters such as carcinoembryonic antigen (CEA), hemoglobin levels, and white-blood-cell count (WBC) for their nomogram. Given the pivotal role of blood and lymph in tumor metastasis, we believe these variables could be essential prognostic features but were, however, absent in the current study because of its retrospective nature. Yang et al. (32) on the other hand, used a renowned scoring system called the Frankel score in their nomogram, which was also not included in the present study. However, this feature might not be predictive of spinal metastasis survival since it was only designed to categorize spinal cord injuries (37).

Access to population-based registries and adding other variables to the nomogram, such as (radi)omics, pathology, and hematological parameters, might further improve the nomograms' performance. Also, accessing these databases will make the nomogram more generalizable by including more primary tumors and increase the number of patients in each primary tumor. At present, the nomogram is limited to five primary tumors, which implies that patients with other primary tumors like cervix, kidney, bladder, etc., cannot benefit from this nomogram.

Conclusions

We have established a user-friendly and easy-to-use prognostic nomogram for patients with SBM using seven known clinical parameters. It has a digital version that can be integrated into the current treatment workflow to aid treatment decision-making in managing cancer patients with SBM. However, proper external validation is needed to ascertain its clinical reliability.

Supplemental materials

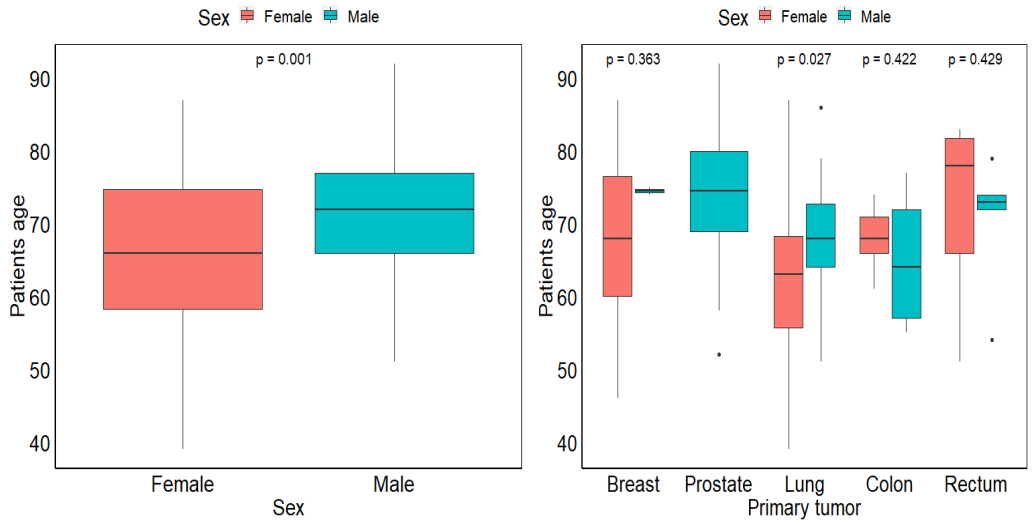

Figure 5.1. *Box-plots showing the age distribution for sex and primary tumor respectively*

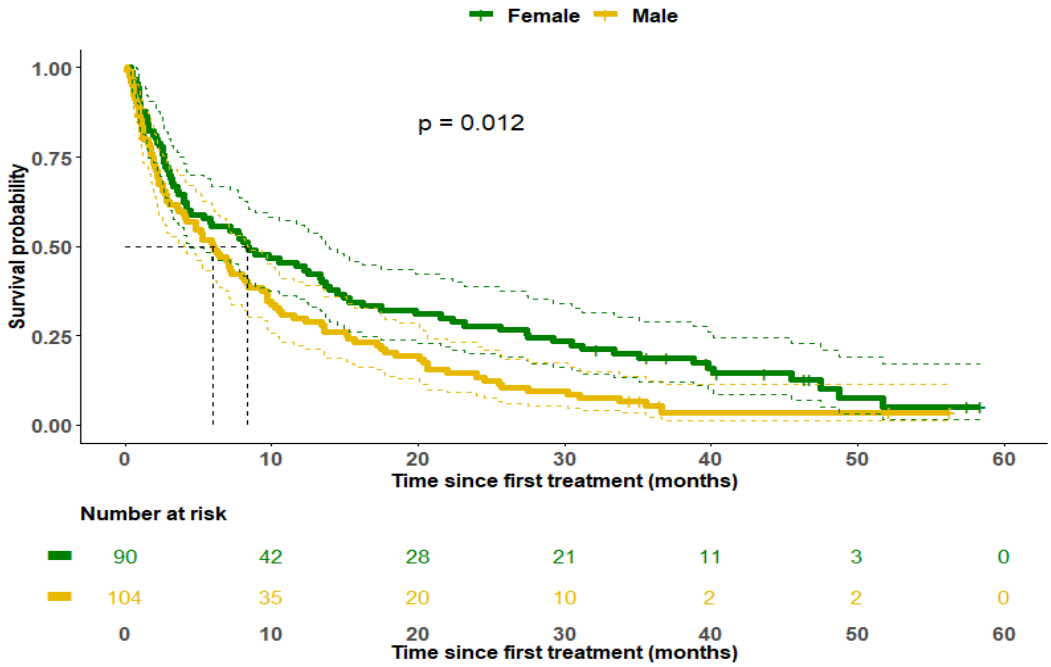

Figure 5.2. *The Kaplan-Meier survival curve for sex.*

Bibliography

[1] Endrit Ziu, Vibhu Krishnan Viswanathan, and Fassil B Mesfin. Spinal metastasis. *StatPearls [Internet]*, 2020.

[2] Hanna Dillekås, Michael S Rogers, and Oddbjørn Straume. Are 90% of deaths from cancer caused by metastases? *Cancer Medicine*, 8(12):5574–5576, 2019.

[3] Lubdha M Shah and Karen L Salzman. Imaging of spinal metastatic disease. *International journal of surgical oncology*, 2011, 2011.

[4] Peter S Rose and Jacob M Buchowski. Metastatic disease in the thoracic and lumbar spine: evaluation and management. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, 19(1):37–48, 2011.

[5] Tse Victor and Arya Kalani Maziyar. Understanding metastatic spine cancer and spinal tumors. https://emedicine.medscape.com/article/1157987-overview. [Accessed: 15 June, 2020].

[6] D. Than Khoi. Understanding metastatic spine cancer and spinal tumors. https://www.spineuniverse.com/resource-center/spinal-cancer/understanding-metastatic-spine-cancer-spinal-tumors. [Accessed: 15 June, 2020].

[7] Amy Yao, Christopher A Sarkiss, Travis R Ladner, and Arthur L Jenkins III. Contemporary spinal oncology treatment paradigms and outcomes for metastatic tumors to the spine: a systematic review of breast, prostate, renal, and lung metastases. *Journal of clinical neuroscience*, 41:11–23, 2017.

[8] Bríain Ó Hartaigh, Heidi Gransar, Tracy Callister, Leslee J Shaw, Joshua Schulman-Marcus, Wijnand J Stuijfzand, Valentina Valenti, Iksung Cho, Jackie Szymonifka, Fay Y Lin, et al. Development and validation of a simple-to-use nomogram for predicting 5-, 10-, and 15-year survival in asymptomatic adults undergoing coronary artery calcium scoring. *JACC: Cardiovascular Imaging*, 11(3):450–458, 2018.

[9] W Van Gijn, Rgpm Van Stiphout, CJH Van De Velde, Vincenzo Valentini, G Lammering, Maria Antonietta Gambacorta, Lars Påhlman, K Bujko, and P Lambin. Nomograms to predict survival and the risk for developing local or distant recurrence in patients with rectal cancer treated with optional short-term radiotherapy. *Annals of Oncology*, 26(5): 928–935, 2015.

[10] Tsuyoshi Konishi, Yoshifumi Shimada, Meier Hsu, Iris H Wei, Emmanouil Pappou, J Joshua Smith, Garrett M Nash, José G Guillem, Philip B Paty, Julio Garcia-Aguilar, et al. Contemporary validation of a nomogram predicting colon cancer recurrence, revealing all-stage improved outcomes. *JNCI cancer spectrum*, 3(2):pkz015, 2019.

[11] Chang-Hsien Lu, Chien-Ting Liu, Pei-Hung Chang, Chia-Yen Hung, Shau-Hsuan Li, Ta-Sen Yeh, Yung-Shin Hung, and Wen-Chi Chou. Develop and validation a nomogram to predict the recurrent probability in patients with major salivary gland cancer. *Journal of Cancer*, 8(12):2247, 2017.

[12] Jun Zhang, Jin Yang, Hai-Qiang Wang, Zhenyu Pan, Xiaoni Yan, Chuanyu Hu, Yuanjie Li, and Jun Lyu. Development and validation of a nomogram for osteosarcoma-specific survival: A population-based study. *Medicine*, 98(23), 2019.

[13] Hang Zhou, Xiong Li, Yuan Zhang, Yao Jia, Ting Hu, Ru Yang, Ke-Cheng Huang, Zhi-Lan Chen, Shao-Shuai Wang, Fang-Xu Tang, et al. Establishing a nomogram for stage ia-iib cervical cancer patients after complete resection. *Asian Pacific journal of cancer prevention: APJCP*, 16(9):3773–3777, 2015.

[14] Tao Wang, Rong Lu, Sunny Lai, Joan H Schiller, Fang Liz Zhou, Bo Ci, Stacy Wang, Xiaohan Gao, Bo Yao, David E Gerber, et al. Development and validation of a nomogram prognostic model for patients with advanced non-small-cell lung cancer. *Cancer informatics*, 18:1176935119837547, 2019.

[15] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

[16] FE Harrell. Regression modeling strategies. as implemented in r package "rms" version 3. *Regression modeling strategies. as implemented in R package 'rms' version*, 3, 2014.

[17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[18] Trevor Hastie and Junyang Qian. Glmnet vignette. *Retrieved June*, 9(2016):1–30, 2014.

[19] Xi Zhou, Haomin Cui, Yu He, Guixing Qiu, Dongsheng Zhou, and Yong Liu. Treatment of spinal metastases with epidural cord compression through corpectomy and reconstruction via the traditional open approach versus the mini-open approach: A multicenter retrospective study. *Journal of oncology*, 2019, 2019.

[20] Yasuaki Tokuhashi, Yasumitsu Ajiro, and Natsuki Umezawa. Outcome of treatment for spinal metastases using scoring system for preoperative evaluation of prognosis. *Spine*, 34(1):69–73, 2009.

[21] Hiroshi Uei and Yasuaki Tokuhashi. Prognostic factors in patients with metastatic spine tumors derived from lung cancer—a novel scoring system for predicting life expectancy. *World journal of surgical oncology*, 16(1):131, 2018.

[22] Henrik CF Bauer and Rikard Wedin. Survival after surgery for spinal and extremity metastases: prognostication in 241 patients. *Acta orthopaedica Scandinavica*, 66(2):143–146, 1995.

[23] Katsuro Tomita, Norio Kawahara, Tadayoshi Kobayashi, Akira Yoshida, Hideki Murakami, and Tomoyuki Akamaru. Surgical strategy for spinal metastases. *Spine*, 26(3): 298–306, 2001.

[24] Yvette M van der Linden, Sander PDS Dijkstra, Ernest JA Vonk, Corrie AM Marijnen, Jan Willem H Leer, and Dutch Bone Metastasis Study Group. Prediction of survival in patients with metastases in the spinal column: results based on a randomized trial of radiotherapy. *Cancer*, 103(2):320–328, 2005.

[25] Panayiotis J Sioutos, Ehud Arbit, Carrie F Meshulam, and Joseph H Galicich. Spinal metastases from solid tumors. analysis of factors affecting survival. *Cancer*, 76(8): 1453–1459, 1995.

[26] H Katagiri, M Takahashi, K Wakai, H Sugiura, T Kataoka, and K Nakanishi. Prognostic factors and a scoring system for patients with skeletal metastasis. *The Journal of Bone and Joint Surgery. British volume*, 87(5):698–703, 2005.

[27] Hirohisa Katagiri, Rieko Okada, Tatsuya Takagi, Mitsuru Takahashi, Hideki Murata, Hideyuki Harada, Tetsuo Nishimura, Hirofumi Asakura, and Hirofumi Ogawa. New prognostic factors and scoring system for patients with skeletal metastasis. *Cancer medicine*, 3(5):1359–1367, 2014.

[28] Andrew J Schoenfeld, Hai V Le, Youssra Marjoua, Dana A Leonard, Philip J Belmont Jr, Christopher M Bono, and Mitchel B Harris. Assessing the utility of a clinical prediction score regarding 30-day morbidity and mortality following metastatic spinal surgery: the new england spinal metastasis score (nesms). *The Spine Journal*, 16(4):482–490, 2016.

[29] Max Aebi. Spinal metastasis in the elderly. In *The Aging Spine*, pages 120–131. Springer, 2005.

[30] M Tevfik Dorak and Ebru Karpuzoglu. Gender differences in cancer susceptibility: an inadequately addressed issue. *Frontiers in genetics*, 3:268, 2012.

[31] P Sutcliffe, M Connock, D Shyangdan, R Court, NB Kandala, and A Clarke. A systematic review of evidence on malignant spinal metastases: natural history and technologies for identifying patients at high risk of vertebral fracture and spinal cord compression. *Health Technology Assessment (Winchester, England)*, 17(42):1, 2013.

[32] Minglei Yang, Wei Xu, Tielong Liu, Xinghai Yang, Peng Wang, Sujia Wu, Haifeng Wei, Jian Zhao, Cheng Yang, and Jianru Xiao. Development and validation of a novel survival prediction model in patients with spinal metastasis from non-small cell lung cancer. *Spine*, 44(4):246–257, 2019.

[33] Yujie Liu, Minglei Yang, Bo Li, Kehan Xu, Xin Gao, Jialin Li, Haifeng Wei, Quan Huang, Wei Xu, and Jianru Xiao. Development of a novel model for predicting survival of patients with spine metastasis from colorectal cancer. *European Spine Journal*, 28(6): 1491–1501, 2019.

[34] Nuno Rui Paulino Pereira, Stein J Janssen, Eva van Dijk, Mitchel B Harris, Francis J Hornicek, Marco L Ferrone, and Joseph H Schwab. Development of a prognostic survival algorithm for patients with metastatic spine disease. *JBJS*, 98(21):1767–1776, 2016.

[35] Tracy L Bale and C Neill Epperson. Sex as a biological variable: who, what, when, why, and how. *Neuropsychopharmacology*, 42(2):386–396, 2017.

[36] Matthew E Arnegard, Lori A Whitten, Chyren Hunter, and Janine Austin Clayton. Sex as a biological variable: a 5-year progress report and call to action. *Journal of Women's Health*, 29(6):858–864, 2020.

[37] Laurens Bollen, Yvette M van der Linden, Willem Pondaag, Marta Fiocco, Bas PM Pattynama, Corrie AM Marijnen, Rob GHH Nelissen, Wilco C Peul, and PD Sander Dijkstra. Prognostic factors associated with survival in patients with symptomatic spinal bone metastases: a retrospective cohort study of 1 043 patients. *Neuro-oncology*, 16(7):991–998, 2014.

Part IV
Decision Tree

CHAPTER 6

PREDICTION OF RADIATION THERAPY COMPLIANCE IN ELDERLY CANCER PATIENTS

Authors:
Biche Osong, *Inigo Bermejo, Kyu Chan Lee, Andre Dekker, Seok Ho Lee, Johan van Soest.*

Adapted from: *Cancers (Special Issue: Personalized Radiation Therapy for Oncology)*
DOI: ***10.3390/cancers14246116***

Abstract

Purpose
This study aims to analyze the relationship between the available variables and treatment compliance in elderly cancer patients treated with radiotherapy and establish a decision tree model to guide caregivers in their decision-making process.

Summary of background
Radiotherapy is one of the most opted forms of cancer treatment. Unfortunately, unwanted interruptions often affect the treatment process, affecting overall survival and local control.

Patients and Methods
For this study purpose, 457 patients over 74 years of age who received radiotherapy between 2005 and 2017 at the Gil medical center in Korea were included in this retrospective analysis. The outcome of interest was radiotherapy compliance, determined by whether patients completed their scheduled radiotherapy treatment (compliance means they completed their treatment and noncompliance otherwise). A bootstrap (B=400) technique was implemented to select the best tunning parameters to establish the decision tree. The developed decision tree's performance in discriminating between compliance and noncompliance patients was evaluated by computing the area under the curve (AUC) and generating calibration plots based on the repeated 5-fold cross-validation method.

Results
The developed decision tree uses patient's status, Charlson comorbidity index, Eastern Cooperative Oncology Group Performance scale, cancer type, age, sex, cancer type, health insurance status, radiotherapy aim, and fractionation type to distinguish between compliance and noncompliance patients. The decision trees' mean AUC and 95% confidence interval was 0.71 (0.66 - 0.77).

Conclusion
We developed and internally validated a novel decision tree that could discriminate between compliance and noncompliance to radiotherapy treatment for elderly cancer patients. The decision tree has a moderate discriminating ability and could serve as a decision aid for caregivers to care better for their elderly cancer patients. However, external validation of the decision tree is warranted to ascertain its clinical usefulness.

Introduction

Approximately 60% of cancer incidence occurs in adults aged 65 and older (1–5). Nowadays, many consider people over 70 years as elderly (4). However, this group of patients, generally called the elderly, are often not granted access to therapeutic clinical trials (6, 7). Therefore, there are still many open questions about the clinical and behavioral responses of elderly cancer patients to cancer treatment (2). Generally, elderly persons are frail due to their relatively weaker immune systems (8). This frailty, combined with common multiple comorbidities, makes them vulnerable to chronic illnesses and even therapy (9, 10), making it arduous to choose the appropriate treatment (2, 11, 12). Although studies have shown that some treatments are feasible for elderly patients (13, 14), the likelihood of their discontinuation is greater than that of the younger generation [15]. This is primarily due to increased comorbidity, decreased performance status, and poor health due to treatment (15).

Radiotherapy, a treatment option that uses ionizing radiation to treat various malignant and benign disorders with curative or palliative intent, is one of the most widely used and effective cancer treatments. However, the treatment process is sometimes not completed as planned, with unwanted interruptions encountered during the treatment process either due to technical reasons or patient-related reasons, such as religious beliefs, financial burden, radiotherapy myths, and travel burden (16–18). Such treatment interruptions may affect local control and overall survival (19) and induce unnecessary treatment-related toxicities for these patients (19–21).

True, some studies have looked at treatment compliance as an outcome of interest either for a particular disease or a combination of diseases (15, 22, 23). However, as far as we are aware, these studies have focused their analysis on a univariate association of the available variables with the endpoint treatment compliance in (elderly) cancer patients. Therefore, a strategy or model to identify elderly cancer patients who might not complete their planned radiotherapy treatment beforehand would be beneficial.

Decision trees, a commonly used prediction technique that can connect several variables to naturally classify patients into various risk groups based on the outcome of interest and present the knowledge graphically to serve as a decision aid (23, 24), are more suitable to model such endpoints.

Therefore, this study aimed to use the available patient information to develop a decision tree that can discriminate between elderly cancer patients based on the likelihood of them completing their planned radiotherapy treatment and use the visual knowledge from the tree to support physicians and caregivers with their decision-making tasks for better patient management.

Materials and methods

After obtaining ethical approval from the internal review board, data for 456 cancer patients above 74 years (elderly) who were treated with radiotherapy were retrospectively collected from the patient's medical records at the Gil Medical Center in Korea between January 2005 and September 2017. All methods were performed following relevant guidelines and regulations.

Radiotherapy was performed using 6 to 15 MV X-ray photons or electron beams with various energies. Compliance, the outcome of interest, was defined as the completion of the prescribed radiotherapy dose, and noncompliance was defined as the discontinuation of therapy by the patient without the caregivers' advice or consent. Only patients with lung, metastatic, head and neck, and gastrointestinal & hepatobiliary cancer were enrolled in this study because they are prevalent cancer types and have a high rate of noncompliance with treatment.

The total radiotherapy dose, fractionation, and radiotherapy field size were determined based on the type of cancer. Patient information such as age, sex, Eastern Cooperative Oncology Group Performance scale (ECOG PS) (25), Charlson comorbidity index (CCI) (26), patient status (in-patient or out-patient), radiotherapy aim, fractionation type, health insurance status, and cancer type was considered for this study. Radiotherapy aim and fractionation type (conventional fractionation and hypofractionation) were considered for treatment information. Only patients treated with conventional fractionation and hypofractionation techniques were included in this study. The stereotactic radiosurgery (SRS) / stereotactic body radiotherapy (SBRT) technique was not included because just one patient was noncompliant with radiotherapy (Figure S6.1 supplemental material). Based on radiotherapy dose and fractionation, patients were classified into either the conventional fractionation (1.8-2.0 Gy fractionation) group, hypofractionation (more than 2 Gy fractionation) group, or stereotactic radiosurgery/ stereotactic body radiotherapy (more than 7.5 Gy fractionation) group.

Health insurance status was used as a surrogate for the actual financial status of the patient. In Korea, the health protection system exempts economically disadvantaged people (patients with medical care insurance) from insurance premiums which means they do not have to pay for their treatment. For indirectly estimating the economic status of elderly patients, medical insurance status was introduced, which can be classified into two patient groups, one with health insurance and the other with medical care in Korea. Patients with health insurance pay part of the treatment fee, and the government pays the rest, whereas, for medical care patients, the government pays the entire treatment cost for the patients.

Decision Trees

A decision tree is a nonlinear discrimination method that uses independent variables to split a sample into progressively smaller subgroups, utilizing binary rules. The basic idea is to recursively partition the covariate space to form subgroups called nodes on the tree for subjects with similar characteristics based on the outcome of interest. The iterative procedure begins with the independent variable with the strongest association with the dependent variable of interest based on specific criteria (27). The first variable or node at the top of the tree is the root node, the tree's most important feature. The other variables that help the tree split further are the internal nodes. Each variable split on the tree is called a branch or edge, and the end of the branch that does not split any further is called the decision or leaf node; in this case, whether a patient completed (compliance) or discontinued (noncompliance) their planned radiotherapy treatment (27, 28).

Statistical analysis

Exploratory analyses and data visualization such as distribution and box plots were applied to gain insights into the data sets and understand the data's underlying patterns. The bootstrap (B = 400) technique and grid search option within the caret package (29) was employed to find the best tuning parameters. The identified optimal hyperparameters were then used to grow the decision tree to predict compliance in elderly cancer patients. The performance of the resulting decision tree was evaluated using a 5-fold cross-validation method. The respective area under the curve (AUC) and calibration for each of the folds were then computed and plotted. Odds ratios at each branch of the tree were computed to measure the association between treatment compliance and the variable. All statistical analyses were performed with the R software (30), and a p-value < 0.05 was considered statistically significant.

Results

A total of 456 elderly cancer patients with a median age of 78 (74-92) years who received radiation therapy were considered to grow and internally validate the decision tree. Table 6.1 shows the demographic and clinical characteristics of the study. There was no difference in the median age or CCI for compliant and noncompliant patients. In contrast, the median fractionation and radiotherapy dose for compliant patients was 3-fold that of noncompliant patients. Metastatic patients had the highest contribution to this study based on sample size and lung cancer patients in terms of noncompliance.

Table 6.1. General patient demographic and clinical characteristics

Variable	Levels	Compliance	Noncompliance	Total
CCI	Median (sd)	6 (1.9)	6 (1.9)	6 (1.9)
Age	Median (sd)	78 (3.5)	78 (3.3)	78 (3.5)
Fractionation	Median (sd)	25 (9.9)	8 (7.2)	25 (11)
RT dose (Gy)	Median (sd)	50.4 (15.4)	16.4 (13.5)	45 (18.9)
Sex	Male	187 (78.6%)	51 (21.4%)	238(52.1%)
	Female	182 (83.1%)	37 (16.9%)	219(47.9%)
ECOG PS	Good (0-2)	322 (79.7%)	82 (20.3%)	404(88.4%)
	Poor (3-4)	47 (88.7%)	6 (11.3%)	53 (11.6%)
Patient Status	In-patient	258 (84.6%)	47 (15.4%)	305(66.7%)
	Out-patient	111 (73.0%)	41 (27.0%)	152(33.3%)
RT Aim	Curative	274 (80.4%)	67 (19.6%)	341(74.6%)
	Palliative	95 (81.9%)	21 (18.1%)	116(25.4%)
Insurance Type	Medical care	50 (86.2%)	8 (13.8%)	58 (12.7%)
	Health insurance	319 (79.9%)	80 (20.1%)	399 (87.3%)
Fraction type	Conventional	242 (82.0%)	53 (18.0%)	295 (64.6%)
	Hypofraction	242 (82.0%)	53 (18.0%)	295 (64.6%)
Cancer Type	Lung	102 (79.7%)	26 (20.3%)	128 (28.0%)
	Metastatic	117 (82.4%)	25 (17.6%)	142 (31.1%)
	Head & Neck	46 (76.7%)	14 (23.3%)	60 (13.1%)
	Gastrointestinal	104 (81.9%)	23 (18.1%)	127 (27.8%)
Total	-	**369 (80.7%)**	**88 (19.3%)**	**789 (100%)**

sd: Standard Deviation, **RT**: Radiotherapy

ECOG PS = Eastern Cooperative Oncology Group Performance Scale.

Biological Effective Dose (BED) (31) and Equivalent Dose in 2 Gy fractions (EQD2) (32) using an α/β ratio 10 are also used to quantify radiobiological concepts into concrete interpretable values.

Focusing only on noncompliant patients (Table 6.1), as it is the group of interest, 87 (19.1%) patients out of the 456 analyzed did not complete their planned radiotherapy treatment (Figure S6.2 supplemental material). More than 50% of the patients in this population did

not complete their treatment due to worsening performance status. The rest of the patients decided not to continue treatment (21%), had radiotherapy-related morbidity (8%), or died during treatment (2%).

Based on table 6.1, men had a higher noncompliant rate than women. Only 6 (11.3%) patients with poor (3-4) ECOG PS did not adhere to their treatment against 81 (20.1%) with good (0-2) ECOG PS. There were similar observations for health insurance status, as there were 79 (19.8%) patients with health insurance but only 8 (13.8%) with medical care. The number of patients treated with curative intent (67 (19.6%)) was 3 times more than that treated with palliative intent (21(18.1%)). The number of noncompliant inpatients was only slightly higher than that of outpatients while those treated with conventional fractionation 53 (18.0%) were more than those treated with hypofractionation 34 (21.1%). With respect to cancer type, lung and metastatic cancer had the highest number of noncompliance 25 (19.7%) and 25 (17.6%) respectively while head and neck cancer had the least 14 (23.3%).

Given that this study is centered around elderly cancer patients, box plots of age and the other considered variables were produced to visualize age distribution within these variables by compliance status. Figure 6.1 shows that there was no significant mean age difference between compliant and noncompliant patients. The same nonsignificant result was observed in all the other variables except for health insurance status and tumor type. Additionally, there was no significant mean age difference within the variables' levels.

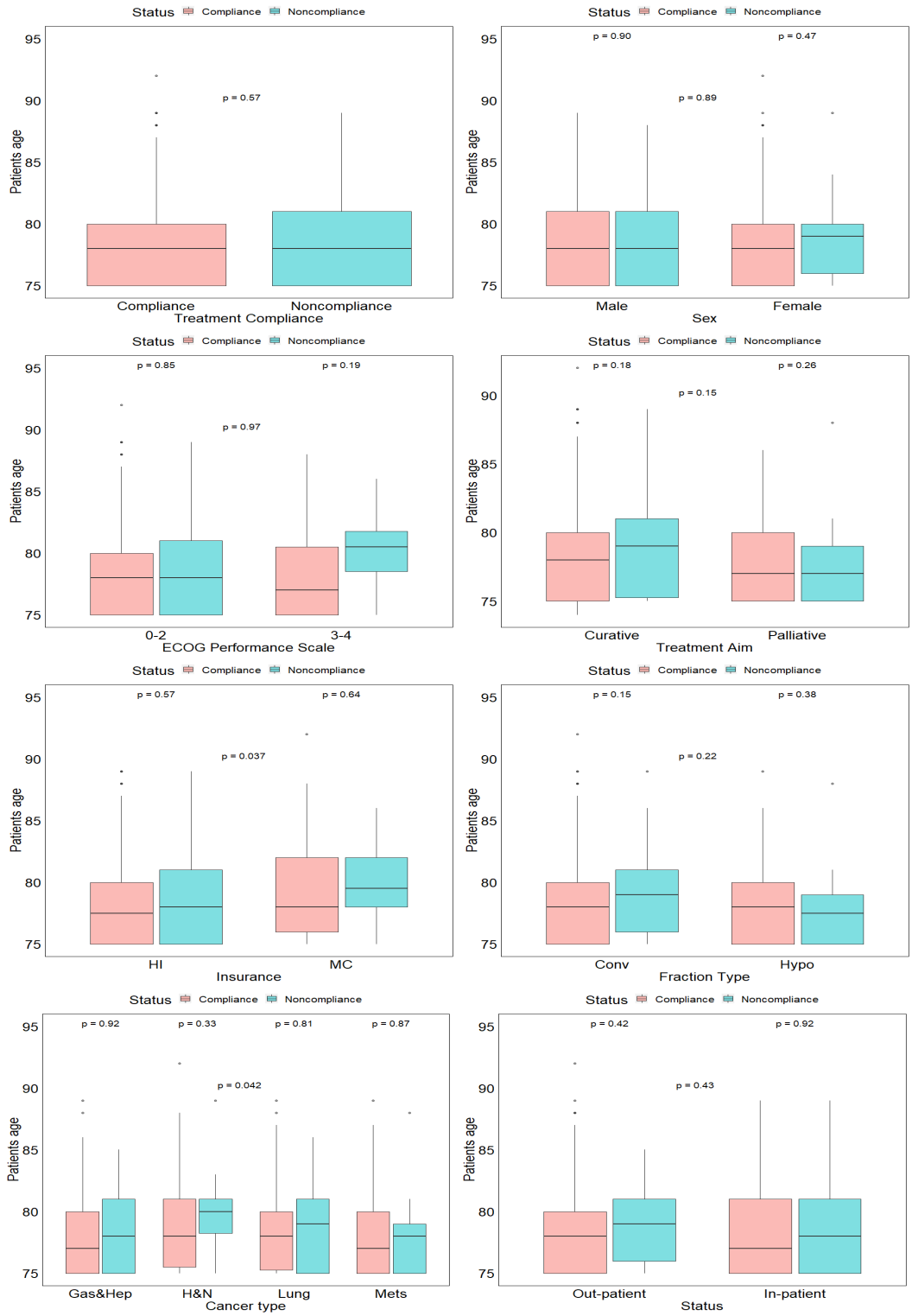

Figure 6.1. *Box-plot of age and all the considered variables by compliance status. p-values in above are within-group comparison and below between-group comparison.*
***HI**: Health insurance, **MC**: Medical care, **Conv**: Conventional fractionation, **Hypo**: Hypofractionation, **GasHep**: Gastrointestinal & Hepatobiliary, **H&N**: Head & Neck, **Mets**: Metastatic, **ECOG**: Eastern Co-operative Oncology Group.*

Decision tree

Based on the bootstrap result (Figure S6.3 supplemental material), a maximum tree depth of 4 was selected with a minimum criterion of 0.041. Figure 6.2 shows the resulting decision tree with these optimal tuning parameters from the bootstrap runs. The oval structures are the independent variables represented as a condition or node, based on which the tree splits into branches or edges. The black text between the thin line on both sides of the nodes describes the split situation that is to be followed to obtain the patient's probability of radiotherapy compliance from the rectangular structures at the bottom of the tree.

To use the decision tree as a decision tool, locate the root node on the top of the tree and read the condition. Then, follow a series of repeated IF-THEN processes based on the patient characteristics on the decision tree until you arrive at the last node, which splits no further. The patients' probability of compliance or noncompliance is then read from the leaf node. On this tree, the most important variable is the patient's status, which splits between inpatients and outpatients. For outpatients, we take the left route, which indicates that the sex of the patient is needed to reach a decision.

In contrast, if the patient is an inpatient, we take the right path, where the second most essential variable for this group of patients needs to be consulted. Here, we checked the patient's CCI. If the patient has a CCI value above 6, we move right, where the patient's insurance status becomes important and split into patients with health insurance and medical care. For medical care patients, their corresponding compliance probability is read directly from the leaf node at the bottom of the tree. In this case, there is an approximately 95% chance that the patient will complete their planned radiotherapy treatment or a 5% chance of noncompliance. On the other hand, if the patient has health insurance, we move left where the treatment aim becomes important, and the tree splits up directly into two leaf nodes, with curative patients having a slightly higher noncompliance probability (25%) compared to palliative patients (20%).

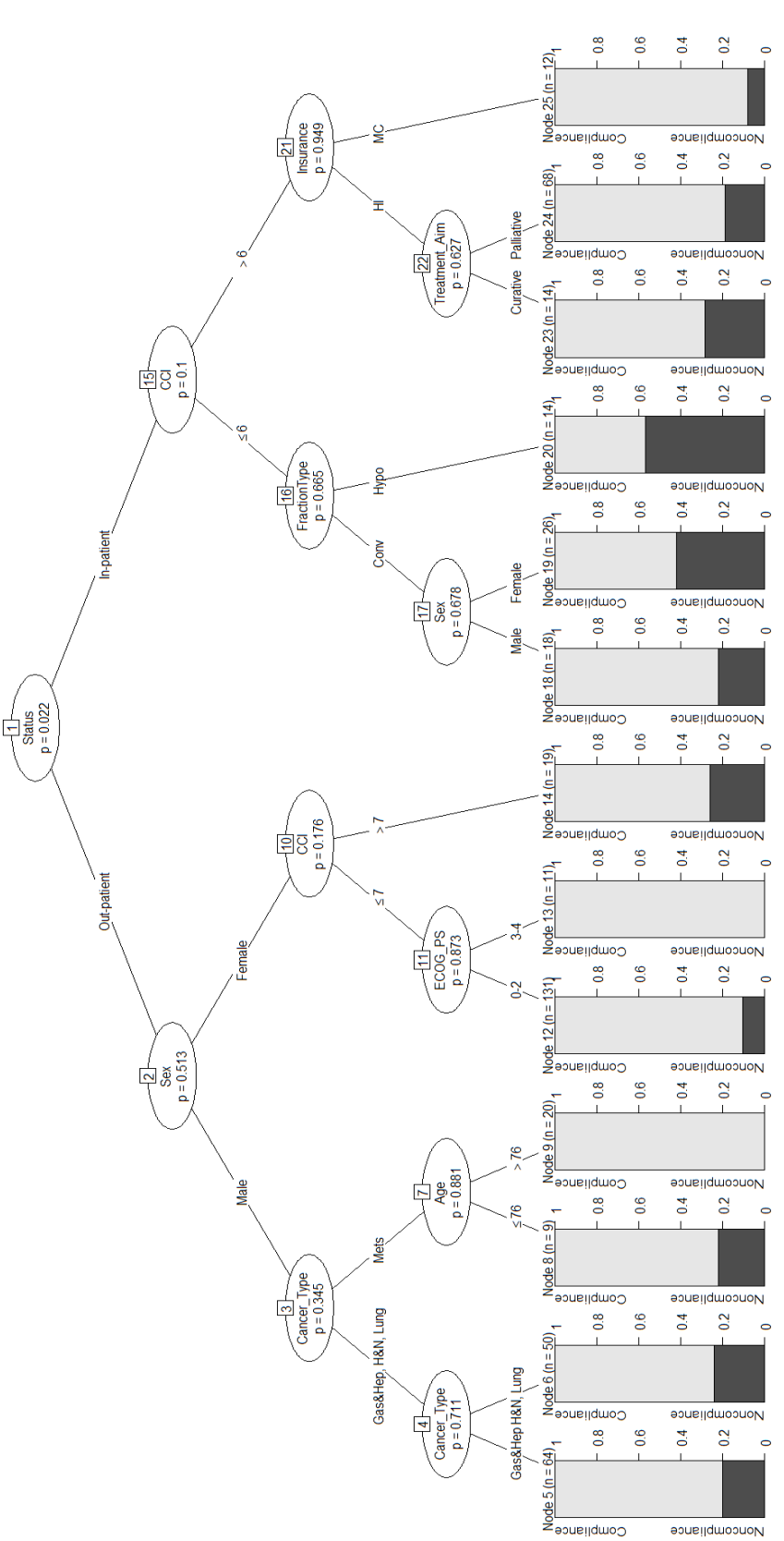

Figure 6.2. *Decision tree for predicting radiotherapy compliance in elderly cancer patients. The oval structures represent the variables that branch out to form the tree. The branches connect the variables and hold the condition for the splits. The rectangle structures which do not branch any further on the tree are the leaf or decision node from which the probability of compliance (white) and noncompliance (black) is read. The values at the top of the leaf nodes indicate the number of patients' in that node, and the p-values in the oval structures indicate the significance level of the variable.*
***HI**: Health insurance, **MC**: Medical care, **H&N**: Head & Neck, **Gas**: Gastrointestinal, **Mets**: Metastatic, **CCI**: Charlson Comorbidity index. **ECOG PS**: Eastern Cooperative Oncology Group Performance Scale.*

Figure 6.3 shows the decision tree's performance for predicting radiotherapy compliance in elderly cancer patients based on the area under the curve (AUC) and the calibration plot. The decision tree's mean AUC and 95% confidence interval (CI) were 0.71 (0.66 - 0.77). The sensitivity and specificity of the developed decision tree were 0.64 and 0.75, respectively, based on a threshold of 0.19. These figures imply that approximately 64% of the patients who completed their planned radiotherapy treatment were correctly classified as patients who adhered to treatment by the tree. On the other hand, 75% of the patients who discontinued their treatment were correctly identified by the tree as noncompliant.

Calibration plots indicate how similar the predicted probabilities are to the actual or observed values. For a perfect or ideal model, all the points should fall precisely on the dotted diagonal gray line. The plot shows good agreement between observed and predicted probabilities for most of the cross-validation samples.

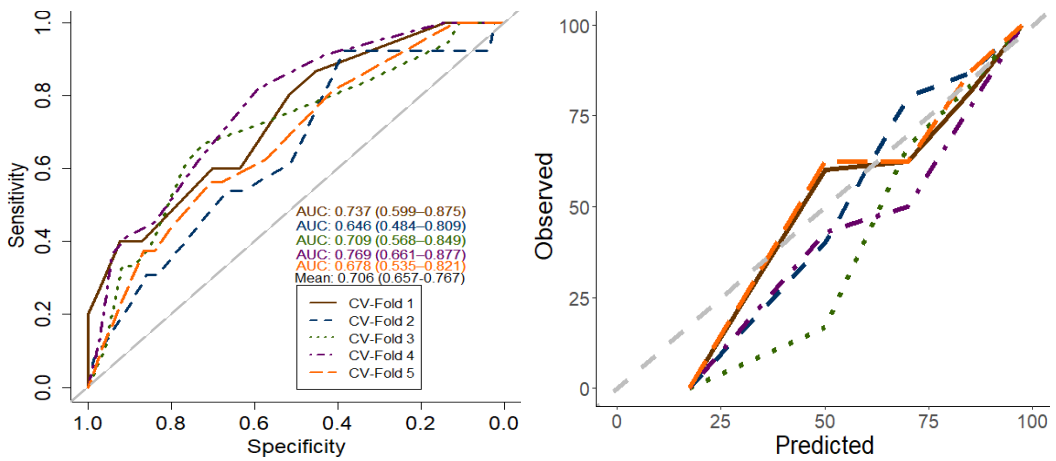

Figure 6.3. *The grown decision tree's performance in terms of the area under the receiver-operating characteristics curve (AUC) and calibration, respectively.*

Discussion

The decision tree methodology was the preferred analysis method because it can simplify complex relationships between the dependent and target variables and make the connections more natural to understand and interpret (24). Unlike other classification models, tree models are more intuitive, self-explanatory, and easy to understand. Decision trees are nonparametric, meaning assumptions or data distributions do not tie them down (23, 33). Therefore, they can be applied to any data to evaluate and account for complex relationships within the data and present the results in a (clinically) usable form (17, 33). Their ability to naturally classify patients into various groups based on the outcome or endpoint of interest makes them a very appealing and handy decision tool in medicine (33).

The current study developed and internally validated a decision tree for predicting radiotherapy compliance in elderly cancer patients. The decision tree had a mean AUC value of 0.71 (0.66 - 0.77), with patient status, sex, cancer type, age, CCI, ECOG PS, fractionation type, treatment aim, and insurance as essential factors to determine radiotherapy compliance. Our findings are similar to Gupta et al. (15) who also analyzed treatment compliance in cancer patients and reported that age, sex, tumor stage, concurrent chemoradiotherapy, and travel distance were significantly associated with noncompliance with radiotherapy. However, it is somewhat challenging to make direct comparisons and interpretations of both studies because of the difference in analytical approaches and study design.

In our study, CCI was a key factor for predicting radiotherapy compliance in elderly patients for both inpatients and outpatients. However, for outpatients, it was only necessary to determine radiotherapy compliance for females. Other studies have also found CCI to be a significant factor affecting adherence to treatments in females, such as chemotherapy and radiotherapy in breast cancer (34). Nonetheless, Di Genesio Pagliuca et al. (13) found no statistically significant correlation between CCI and chemotherapy in a mixed population of 137 (60% males and 40% females) elderly patients treated with chemoradiotherapy.

These contradictory results on the importance of the CCI to predict treatment compliance can be attributed to the difference in the patient population under consideration. The decision tree shows that CCI is a pivotal factor in treatment compliance, but for certain patient groups. Therefore, without proper subgroup analysis or methods such as decision trees, important predictor variables could be easily missed, leading to contradictory results. In addition, the study did report that most of the patients who stopped chemoradiotherapy had a relatively higher CCI and poor performance status. In this study, the odds of noncompliance for an outpatient female with a CCI below 7 are just 0.25 times the odds of an outpatient female with a CCI value above 7. While the noncompliance odds for inpatients with a CCI value below 6 was 2.55 times the odds of noncompliance for inpatients

with a CCI value above 6 (Table S6.1 supplemental material) and all the odds ratios were statistically significant.

This study could not include information about the morbidity of elderly cancer patients because of too much missing information. Hence, we could not evaluate the relationship between morbidities and radiotherapy compliance. Nevertheless, we postulate that good and poor ECOG PS patients could not complete their scheduled radiotherapy due to treatment-related morbidities. However, patients with poor ECOG PS would be more affected by radiotherapy treatment than those with good ECOG PS. In the present study, patients with poor ECOG PS had a lower median age in the compliance group than in the noncompliance group (Figure 6.1). This age difference was absent for patients with good ECOG PS. Based on the developed decision tree, the ECOG PS is only pivotal for outpatient females with a CCI value of less than 7. Patients with good ECOG PS have a relatively higher noncompliance probability than those with poor ECOG PS.

In general, except for the ECOG PS before radiotherapy, numerous factors can affect a patient's ECOG PS during treatment. These include oral mucositis and esophagitis, both of which are typical side effects of radiotherapy either for head and neck cancer or lung cancer. In the study by Yoon et al. (35), they reported that the main reasons for discontinuation of radiotherapy treatment in elderly lung cancer patients were that five patients (42%) were experiencing aggravation of the general condition and cancer progression, and seven patients (58%) were experiencing treatment-related toxicity. They concluded that physicians should pay attention to selecting elderly cancer patients and chemotherapy agents considering general conditions and toxicity before planning concurrent chemoradiotherapy.

In terms of indirectly estimating the economic status of elderly patients, we analyzed radiotherapy compliance according to medical insurance status, classified into patients with health insurance and patients with medical care. As per our institution's policy, insured patients who decide to receive radiotherapy should pay the treatment fee at every visit and not the total cost at the initial appointment. Generally, the radiotherapy treatment fee is higher than the other treatment modalities, which means that cancer patients need for radiotherapy funds is still substantial relative to the other treatments and can cause significant "financial distress." Therefore, patients with health insurance are more likely to have a financial burden than patients with medical care since the government covers the treatment costs. As a result, the noncompliance rate in the health insurance group was tenfold higher than that in patients with medical care.

Although there was a statistically significant age difference between these two insurance groups, noncompliant patients had a higher median age within each group, especially for medical care patients (Figure 6.1). Based on the decision tree, inpatients with a CCI value greater than six and medical care have a higher compliance probability than patients with

health insurance. Also, at this node (node 21), the odds of noncompliance for an inpatient with a CCI above 6 is reduced by approximately 60% if the patient has medical care insurance (Table S6.1 supplemental material). For health insurance patients, compliance is mostly affected by the type of treatment they receive, with patients treated with curative intent having a relatively higher noncompliance probability than palliative patients (Figure 6.2). More so the odds of noncompliance for patients who received palliative care (node 22) treatment was 0.58 times the odds of noncompliance for curative treatment patients (Table S6.1 supplemental material).

The developed decision tree results can help caregivers and physicians in their decision-making process. For instance, an 80-year-old male not admitted to the hospital (outpatient) and suffering from a metastatic disease is expected to complete his planned radiotherapy schedule based on the decision tree. In such a case, the physician can recommend the most appropriate radiotherapy technique for the patient regardless of age by considering the feasibility of completing treatment based on objective results from the decision tree. On the other hand, an outpatient female with a CCI value above 7 has a 25% chance of not completing her planned radiotherapy treatment. Such information can assist in selecting the most appropriate treatment for this patient or channeling the right resources for a better outcome.

Generally, when radiotherapy treatment lasts longer than seven weeks, it induces fatigue in patients. This is especially true for the elderly, particularly as their physical and mental status is impacted. Hence, based on the decision tree, physicians could propose a shorter radiotherapy treatment period technique, such as stereotactic body radiotherapy (SBRT), for patients predicted to be noncompliant with radiotherapy treatment.

In a study by Amini et al. (36), the authors used SBRT with a relatively short duration of treatment compared to conventional split radiotherapy for elderly head and neck cancer patients with poor performance status. Hayashi et al. (37) also reported the effects of SBRT in elderly patients with stage I non-small-cell lung cancer (NSCLC). They concluded that SBRT for stage NSCLC was well tolerated and feasible in very elderly patients, although elderly patients experienced significantly more severe radiation pneumonitis.

Recent advances in radiotherapy techniques that reduce radiation-related adverse effects, such as intensity-modulated radiation therapy, SBRT, and stereotactic radiosurgery, should be considered for treating elderly cancer patients. Although we did not enroll SRS/SBRT-treated patients in this study, we did visualize the distribution of compliant and noncompliant patients within the different fractionation types, which shows that all but one patient was compliant with radiotherapy treatment as expected (Figure S6.1 supplemental material).

In this study, the number of outpatients who discontinued their planned radiotherapy treatment was higher than inpatients, with a significant difference in the odds of noncompliance (OR = 0.48, 95% CI [0.29 - 0.78]) between the two patient status groups (Table S6.1 supplemental material). This outpatient noncompliance majority was more pronounced in the lung and gastrointestinal/hepatobiliary cancer patients (Figure S6.4 supplemental material). Extended fractionation schedules will require outpatients to commute between the radiotherapy center and their residence. Since elderly patients are more susceptible to treatment-related fatigue and deficits in physical activity, these constant travels can severely impact their quality of life, making it infeasible for them to come for future treatments.

To increase adherence to planned radiotherapy treatment in this patient group, Palwe et al. (18) suggested providing accommodation for these patients closer to the treatment center or having more frequent outpatient visits after the third week of planned radiotherapy treatment. Contrary to these two cancer types, the number of noncompliant outpatients with metastatic or head & neck cancer was lower or equal to inpatients respectively (Figure S6.4 supplemental material). A possible explanation for this difference could be that the general condition of patients in this group is relatively poor and can deteriorate easily or the radiation treatment was administered basically to relieve severe metastasis-related symptoms such as pain and neurologic symptoms. This increases the chance of a patient discontinuing treatment if symptoms were immediately alleviated during treatment or could be advised by a caregiver if the patient's condition worsens.

To the best of our knowledge, this study is the first to assess the predictive value of decision trees for radiotherapy compliance in elderly cancer patients. Therefore, there is room for improvement. As a retrospective study derived from a single institution, this analysis could not include other information, such as morbidity, during radiotherapy. Therefore, external validation of the developed decision tree is needed to ascertain its clinical usefulness.

Nonetheless, the performance of the developed decision tree to determine if a patient will complete their planned radiotherapy treatment is better than toasting a coin. In addition, 75% of the patients who discontinued their treatment was correctly identified by the tree as noncompliant, which provides caregivers with a subgroup of patients to monitor very closely to help prevent them from discontinuing their treatment since it might lead to an unnecessary increase in treatment cost and time wasted for both the patient and caregiver.

Conclusion

In conclusion, we have developed and internally validated a decision tree to predict radiotherapy compliance in elderly cancer patients. Based on the decision tree, treatment compliance mainly depends on the patient's status. Other clinical, social, demographic, and treatment features, such as CCI, ECOG PS, age, sex, cancer type, health insurance status, radiotherapy aim, and fractionation type, also influenced treatment compliance. The developed tree has a reasonably good ability to identify those patients who are likely to discontinue their radiotherapy treatment, giving caregivers a better rationale to decide whether to start a radiotherapy treatment or look for alternative treatment for these patients. The developed decision tree has a moderate discriminating ability and could serve as an aid for caregivers to select the optimal treatment for elderly cancer patients although external validation is needed to determine its clinical usefulness. And, the developed decision tree can also help boost treatment compliance by targeting those patients who are likely to discontinue therapy with incentives and techniques to support them adhere to treatment, especially for patients already receiving therapy.

Supplemental materials

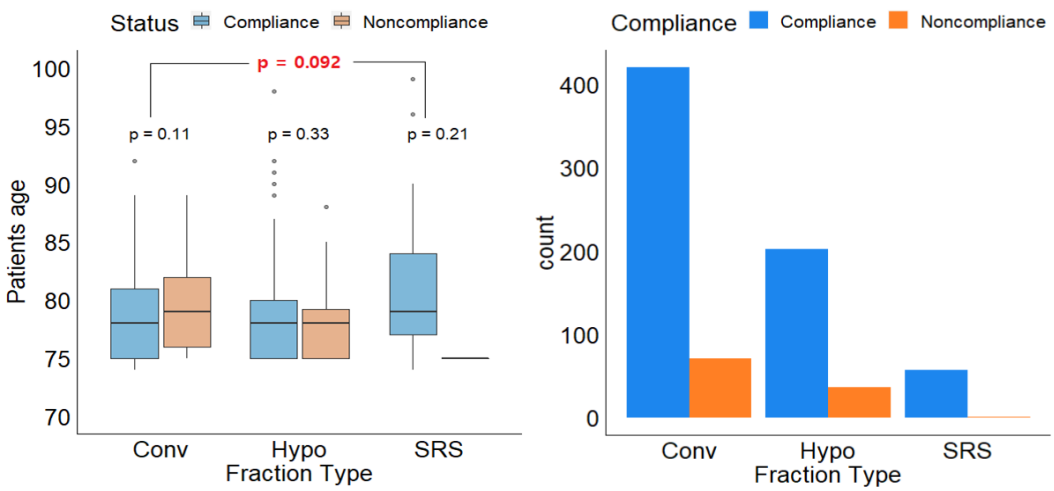

Figure 6.1. *Box and Bar plots for fractionation type by compliance status*

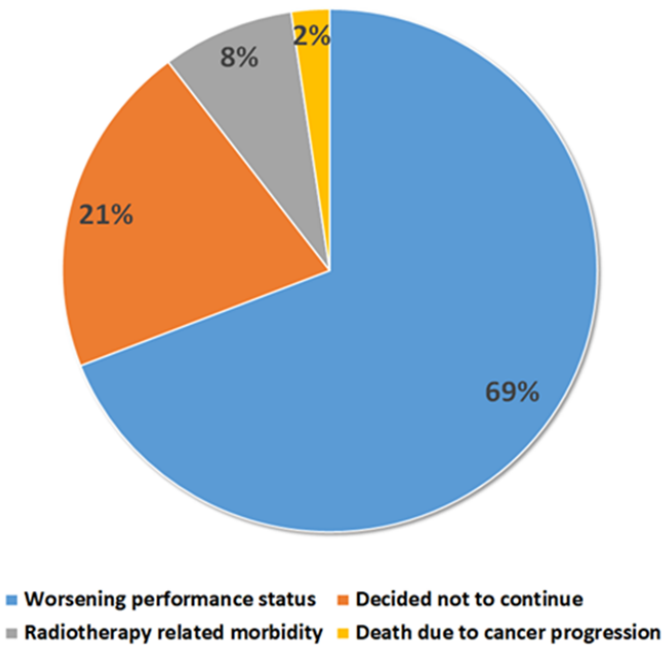

Figure 6.2. *Reasons for radiotherapy noncompliance*

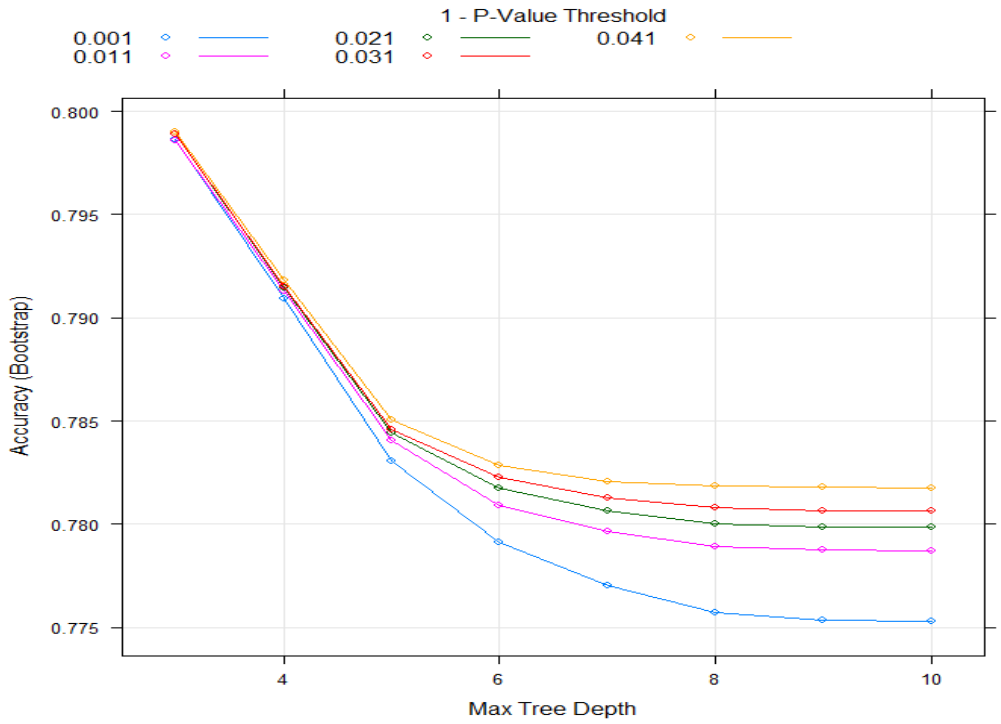

Figure 6.3. *Bootstrap output for selecting optimal parameters*

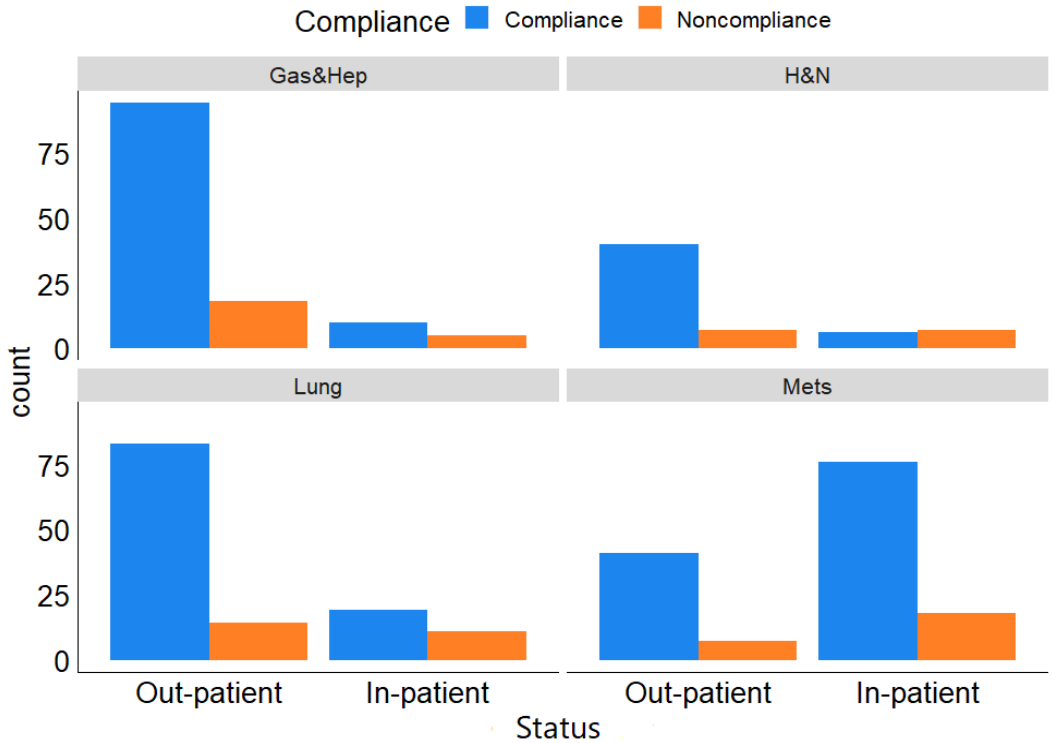

Figure 6.4. *Bar plots of compliance by patients status for each disease.*

Table 6.1. The odds ratio of noncompliance with the corresponding 95% confidence interval and p-value for each branch on the developed decision tree to predict compliance in elderly cancer patients.

Node	Variable	Levels	Nonc	Comp	OR (95% CI)	P-value
1	Patient Status	Out-patient	46	258	0.48 (0.29 - 0.78)	$<$0.05
		In-patient	41	111		
2	Sex	Female	19	142	0.57 (0.30 - 1.08)	0.08
		Male	27	116		
3	Cancer Type	Metastatic	02	27	0.28 (0.04 - 1.04)	0.06
		Others	25	89		
4	Cancer Type	Lung	11	38	1.05 (0.42 - 2.60)	0.90
		Others	14	51		
7	Patient Age	Above 76	00	20	-	-
		Below 76	02	07		
10	CCI	Below 7	14	128	0.25 (0.09 – 0.91)	0.03
		Above 7	05	14		
11	ECOG-PS	3-4	00	11	–	-
		0-2	14	117		
15	CCI	Below 6	23	35	2.55 (1.32 – 5.67)	$<$0.05
		Above 6	18	76		
16	Fraction Type	Conventional	15	29	0.39 (0.10 - 1.37)	0.12
		Hypofraction	18	06		
17	Sex	Female	11	15	2.47 (0.65 - 11.0)	0.18
		Male	15	29		
21	Insurance	MC	01	11	0.39 (0.01 - 2.28)	0.31
		HI	17	65		
22	Treatment aim	Palliative	13	55	0.58 (0.16 - 2.49)	0.43
		Curative	04	10		

Nonc: Noncompliance, **Comp**: Compliance, **OR**: Odds ratio, **CI**: Confidence Interval
CCI: Charlson comorbidity index, **MC**: Medical Care, **HI**: Health Insurance
ECOG-PS: Eastern Cooperative Oncology Group Performance Scale

Bibliography

[1] Jill A Foster, Gregory D Salinas, Dorcas Mansell, James C Williamson, and Linda L Casebeer. How does older age influence oncologists' cancer management? *The Oncologist*, 15(6):584–592, 2010.

[2] Barbara Given and Charles W Given. Older adults and cancer treatment. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 113(S12):3505–3511, 2008.

[3] Seri Hong, Young-Joo Won, Young Ran Park, Kyu-Won Jung, Hyun-Joo Kong, Eun Sook Lee, et al. Cancer statistics in korea: incidence, mortality, survival, and prevalence in 2017. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, 52(2): 335, 2020.

[4] Yasuyoshi Ouchi, Hiromi Rakugi, Hidenori Arai, Masahiro Akishita, Hideki Ito, Kenji Toba, Ichiro Kai, et al. Redefining the elderly as aged 75 years and older: Proposal from the joint committee of japan gerontological society and the japan geriatrics society. *Geriatr Gerontol Int*, 17(7):1045–1047, 2017.

[5] Jigisha P Thakkar, John L Villano, and Bridget J McCarthy. Age-specific cancer incidence rates increase through the oldest age groups. *The American journal of the medical sciences*, 348(1):65–70, 2014.

[6] NA De Glas, M Kiderlen, AJM de Craen, ME Hamaker, JEA Portielje, CJH van de Velde, GJ Liefers, and E Bastiaannet. Assessing treatment effects in older breast cancer patients: systematic review of observational research methods. *Cancer Treatment Reviews*, 41(3): 254–261, 2015.

[7] L Repetto, A Venturino, L Fratino, D Serraino, G Troisi, W Gianni, and M Pietropaolo. Geriatric oncology: a clinical approach to the older patient with cancer. *European Journal of Cancer*, 39(7):870–880, 2003.

[8] Jean-Philippe Gouin, Liisa Hantsoo, and Janice K Kiecolt-Glaser. Immune dysregulation and chronic stress among older adults: a review. *Neuroimmunomodulation*, 15(4-6): 251–259, 2008.

[9] Cecilia G Ethun, Mehmet A Bilen, Ashesh B Jani, Shishir K Maithel, Kenneth Ogan, and Viraj A Master. Frailty and cancer: implications for oncology surgery, medical oncology, and radiation oncology. *CA: a cancer journal for clinicians*, 67(5):362–377, 2017.

[10] Ankush Gosain and Luisa A DiPietro. Aging and wound healing. *World journal of surgery*, 28(3):321–326, 2004.

[11] Alessandra Marengoni, Sara Angleman, René Melis, Francesca Mangialasche, Anita Karp, Annika Garmen, Bettina Meinow, and Laura Fratiglioni. Aging with multimorbidity: a systematic review of the literature. *Ageing research reviews*, 10(4):430–439, 2011.

[12] Alessandro Nobili, Silvio Garattini, and Pier Mannuccio Mannucci. Multiple diseases and polypharmacy in the elderly: challenges for the internist of the third millennium. *Journal of comorbidity*, 1(1):28–44, 2011.

[13] M Di Genesio Pagliuca, C Perotti, G Apicella, A Galla, M Guffi, M Paolini, L Donis, V Amisano, S Torrente, I Manfredda, et al. Concurrent chemo-radiotherapy in elderly patients: tolerance and compliance in a series of 137 patients. *Clinical and Translational Oncology*, 18(6):571–575, 2016.

[14] Erik Haehl, Alexander Rühle, Hélène David, Tobias Kalckreuth, Tanja Sprave, Raluca Stoian, Christoph Becker, Andreas Knopf, Anca-Ligia Grosu, and Nils H Nicolay. Radiotherapy for geriatric head-and-neck cancer patients: what is the value of standard treatment in the elderly? *Radiation Oncology*, 15(1):1–12, 2020.

[15] Shivani Gupta, Kartick Rastogi, Aseem Rai Bhatnagar, Daleep Singh, Kampra Gupta, Ajay Singh Choudhary, et al. Compliance to radiotherapy: A tertiary care center experience. *Indian Journal of cancer*, 55(2):166, 2018.

[16] Ms Mounica Vallurupalli, Ms Katharine Lauderdale, Michael J Balboni, Andrea C Phelps, Susan D Block, Andrea K Ng, Lisa A Kachnic, Tyler J VanderWeele, and Tracy A Balboni. The role of spirituality and religious coping in the quality of life of patients with advanced cancer receiving palliative radiation therapy. *The journal of supportive oncology*, 10(2):81, 2012.

[17] Erin E Kent, Laura P Forsythe, K Robin Yabroff, Kathryn E Weaver, Janet S de Moor, Juan L Rodriguez, and Julia H Rowland. Are survivors who report cancer-related financial problems more likely to forgo or delay medical care? *Cancer*, 119(20):3710–3717, 2013.

[18] Vijay Palwe, Roshankumar Patil, Prakash Pandit, and Rajnish Nagarkar. Factors influencing non-adherence to radiotherapy: a retrospective audit of 1,548 patients from a tertiary cancer centre. *Journal of Radiotherapy in Practice*, 19(4):359–364, 2020.

[19] José A González Ferreira, Javier Jaén Olasolo, Ignacio Azinovic, and Branislav Jeremic. Effect of radiotherapy delay in overall treatment time on local control and survival in head and neck cancer: review of the literature. *Reports of Practical Oncology and Radiotherapy*, 20(5):328–339, 2015.

[20] Branislav Jeremic, Yuta Shibamoto, Biljana Milicic, Aleksandar Dagovic, Nebojsa Nikolic, Jasna Aleksandrovic, Ljubisa Acimovic, and Slobodan Milisavljevic. Impact of treatment interruptions due to toxicity on outcome of patients with early stage (i/ii) non-small-cell lung cancer (nsclc) treated with hyperfractionated radiation therapy alone. *Lung Cancer*, 40(3):317–323, 2003.

[21] Kimberly Thomas, Travis Martin, Ang Gao, Chul Ahn, Holly Wilhelm, and David L Schwartz. Interruptions of head and neck radiotherapy across insured and indigent patient populations. *Journal of oncology practice*, 13(4):e319–e328, 2017.

[22] Marianna Christodoulou, Fiona Blackhall, Hitesh Mistry, Ahmet Leylek, Joost Knegjens, Vincent Remouchamps, Isabelle Martel-Lafay, Núria Farré, Matjaž Zwitter, Delphine Lerouge, et al. Compliance and outcome of elderly patients treated in the concurrent once-daily versus twice-daily radiotherapy (convert) trial. *Journal of Thoracic Oncology*, 14(1):63–71, 2019.

[23] Alfredo Echeverria, Haley Manley, Barrett O'Donnell, Joshua Asper, Mark Bonnen, and Michelle Ludwig. Factors associated with radiation treatment compliance for women with cervical cancer in a safety net health system. *International Journal of Gynecologic Cancer*, 27(7), 2017.

[24] Yan-Yan Song and LU Ying. *Decision tree methods: applications for classification and prediction*, volume 27. Shanghai Mental Health Center, 2015.

[25] Martin M Oken, Richard H Creech, Douglass C Tormey, John Horton, Thomas E Davis, Eleanor T McFadden, and Paul P Carbone. Toxicity and response criteria of the eastern cooperative oncology group. *American journal of clinical oncology*, 5(6):649–656, 1982.

[26] Mary E Charlson, Peter Pompei, Kathy L Ales, and C Ronald MacKenzie. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*, 40(5):373–383, 1987.

[27] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.

[28] Gordon S Linoff and Michael JA Berry. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 2011.

[29] Max Kuhn, Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, R Core Team, et al. Package 'caret'. *The R Journal*, 223:7, 2020.

[30] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[31] Jack F Fowler. Brief summary of radiobiological principles in fractionated radiotherapy. In *Seminars in Radiation Oncology*, volume 2, pages 16–21. Elsevier, 1992.

[32] GG Steel. Basic clinical radiobiology. london, new york, 1997.

[33] Chen Hu and Jon Arni Steingrimsson. Personalized risk prediction in clinical oncology research: applications and practical issues using survival trees and random forests. *Journal of biopharmaceutical statistics*, 28(2):333–349, 2018.

[34] Daniel Wollschläger, Xiaoyu Meng, Achim Wöckel, Wolfgang Janni, Rolf Kreienberg, Maria Blettner, and Lukas Schwentner. Comorbidity-dependent adherence to guidelines and survival in breast cancer—is there a role for guideline adherence in comorbid breast cancer patients? a retrospective cohort study with 2137 patients. *The breast journal*, 24(2):120–127, 2018.

[35] Won-Sup Yoon, Dae-Sik Yang, and Chul-Yong Kim. The evaluation of radiation therapy and combined-modality therapy for non-small-cell lung cancer in elderly. *Radiation Oncology Journal*, 25(2):101–108, 2007.

[36] Arya Amini, Jessica D McDermott, Gregory Gan, Shilpa Bhatia, Whitney Sumner, Christine M Fisher, Antonio Jimeno, Daniel W Bowles, David Raben, and Sana D Karam. Stereotactic body radiotherapy as primary therapy for head and neck cancer in the elderly or patients with poor performance. *Frontiers in Oncology*, 4:274, 2014.

[37] Shinya Hayashi, Hidekazu Tanaka, Yuuichi Kajiura, Yasushi Ohno, and Hiroaki Hoshi. Stereotactic body radiotherapy for very elderly patients (age, greater than or equal to 85 years) with stage i non-small cell lung cancer. *Radiation Oncology*, 9(1):1–9, 2014.

CHAPTER 7

OVERALL SURVIVAL PREDICTION FOR STAGE IIB-IVA CERVICAL CANCER PATIENTS

***Authors**:*
*Seok Ho Lee, **Biche Osong**, Ludy Lutgens, Kwang Beom Lee, Leonard Wee, Andre Dekker, Johan van Soest*

***Under review**:*

Abstract

Purpose
This study aims to develop and validate a decision tree algorithm to predict the overall survival of IIB-IVA cervical cancer patients with squamous cell carcinomas.

Summary of background
Cervical cancer claims approximately 200 lives of women in the Netherlands and over 800 in Korea each year. In less developed countries with limited healthcare services, cervical cancer-related deaths are increasing. Knowing expected survival in an individual patient may help direct the right resources to treat this patient.

Patients and Methods
A total of 146 women diagnosed with cervical cancer and treated between 2000 to 2017 at Maastro and Gil Medical center were analyzed to develop and validate the decision tree. We used the rpart package in R to grow the decision tree and the concordance index (C-index) to measure its discriminating ability.

Results
The developed decision tree requires only three features to predict the overall survival of these patients. The most important predictor for all patients was their FIGO stage. Age at radiotherapy and the SCC-Ag level were the second and third most important features for FIGO stage IIB patients. While for patients with FIGO stage III-IVA tumor, tumor size and the SCC-Ag level were the most important variables in predicting their overall survival. The decision tree's performance on the training cohort reached a C-index of 0.786 (95% CI, 0.711 - 0.860) and 0.716 (95% CI, 0.639 - 0.893) on the validation cohort.

Conclusion
We have developed and externally validated a decision tree to predict the overall survival probability for cervical cancer patients. The tree separates patients into different risk groups based on their survival probabilities which could be leveraged to manage cervical cancer patients better.

Introduction

Thanks to early screening, the mortality rate of cervical cancer has dramatically decreased over the last decades (1). Regardless of all these endeavors, cervical cancer remains the second most common form of cancer-related death in women worldwide (2, 3). Numerically, over 25 000 women die from this disease in Europe each year, and over 200 women in the Netherlands, specifically (4–6). In Korea, about 800 cervical cancer-related deaths occur each year (7). These numbers are even higher in less-developed countries and areas with limited healthcare services (3, 8).

Besides health care accessibility, numerous risk factors are associated with these death rates. The International Federation of Gynecologists and Obstetricians (FIGO) score is a prominent prognostic factor for estimating survival for cervical cancer patients. However, several studies have shown that other clinical and pathological information could improve patient survival prediction models (9–12). These survival models can be graphically represented to serve as decision aid tools such as nomograms and decision trees, where a patient's overall survival (OS) probability can be determined from these clinical and pathological factors (12).

Prediction models such as decision trees are gradually being valued in the medical sector because of their high explainability and interpretability. Their ability to naturally classify patients into various risk groups based on the outcome of interest makes them a handy decision tool (13–15). Until this study's analysis, to the best of our knowledge, no decision tree model existed to predict the overall likelihood of surviving cervical cancer, which is essential for treatment modality decision-making.

This study, therefore, aims at developing and externally validating a decision tree model to predict the overall survival (OS) for locally advanced (IIB-IVA) cervical cancer patients with squamous cell carcinomas.

Materials and methods

A total of 76 patients diagnosed with cervical cancer and treated from 2000 to 2010 at Maastro clinic, Maastricht, The Netherlands, were extracted to train the decision tree. We used an external set of 70 patients from Gil medical center Korea to validate the decision tree. Patients with a histological type of cervical cancer different from squamous, FIGO stage less than IIB, or treated with hyperthermia were excluded from the analyses. We focused on locally advanced cervical cancer patients who are treated with definitive platinum-based concurrent chemoradiation therapy (CCRT). Extracted parameters common to both treatment institutions include the FIGO stage (IIB, IIIA, IIIB, and IVA), patient age at radiotherapy (in years), tumor size, squamous cell carcinoma antigen (SCC-Ag) level, and clinical lymph node involvement based on clinical examination and abdominopelvic CT and/or pelvis MRI imaging. The variable of interest is the overall survival (OS), defined as the time difference between tumor diagnosis and death or last follow-up. The tumor size was determined by measuring the central tumor diameter on abdominopelvic CT and/or pelvis MRI.

Statistical analysis

The preliminary analysis explored the data with some descriptive statistics and graphs. Tumor size was binarized into patients with a tumor size above and below 4cm and excluded patients with missing information from the tree-building process. The variable selection procedure was incorporated into the tree-growing process since the algorithm can select essential variables based on the response. Survival risk groups were constructed from the decision tree splitting leaf nodes and visualized with Kaplan–Meier plots. We compared survival between the risk groups using the log-rank test. The rpart package with default parameter settings was used to construct the decision tree. The concordance index (C-index), which reflects the accuracy of survival models, was used to evaluate the tree's performance. A C-index value of 1 indicates a perfect prediction, while a C-index value of 0.5 is comparable to a random guess (16). All statistical analyses were performed with the R software (17), and a p-value < 0.05 was considered statistically significant.

Results

A total of 146 women with cervical cancer were considered to grow and validate the decision tree. Excluded patients with missing information reduced the training and validation cohorts to 68 and 69, respectively. The median age of patients at Maastro was 59 (34-89) and 61 (26-84) at Gil medical center. We found no statistical difference between the two institutions regarding age and FIGO stage. The median follow-up time in months at Maastro and Gil medical center was 58.51 (49.74-87.36) and 46.87 (21.40-66.93), respectively. Table 7.1 shows the patient and tumor characteristics from the two institutions.

Table 7.1. General patient characteristics at Gil medical center and Maastro

Variable	Levels	Gil MC	Maastro	p-value
Age	Mean (sd)	61 (12.51)	59 (13.20)	0.142
SCC-Ag	Mean (sd)	19 (22.25)	12 (15.22)	0.028
	Missing	01 (1.67%)	08 (10.53%)	
FIGO	IIB	46 (65.71%)	40 (58.82%)	0.779
	IIIA	05 (07.14%)	04 (05.88%)	
	IIIB	14 (20.00%)	18 (26.47%)	
	IVA	05 (07.14%)	06 (08.82%)	
Tumor size	$< 4cm$	10 (14.29%)	13 (19.12%)	< 0.05
	$\geq$ 4cm	60 (85.71%)	55 (80.88%)	
	Missing	00 (00.00%)	01 (01.32%)	
Clinical N	Normal	21 (30.00%)	69 (91.18%)	< 0.05
	Metastases	49 (70.00%)	06 (8.82%)	

FIGO: The International Federation of Gynecology and Obstetrics
SCC-Ag: Squamous Cell Carcinoma Antigen, **MC:** Medical Center , **sd:** Standard Deviation.

The mean patient age at the Gil medical center was slightly higher than at Maastro, with fewer patients having a pelvis nodal metastasis and lower mean SCC-Ag level. A more favorable survival was found in the Gil medical center, as depicted in Figure 7.1. Based on the FIGO staging, stage IIIA, and lower patients were about 72.8% in Gil medical center. In contrast, it was 64.7% for Maastro, with more patients having a FIGO stage IIIB or higher.

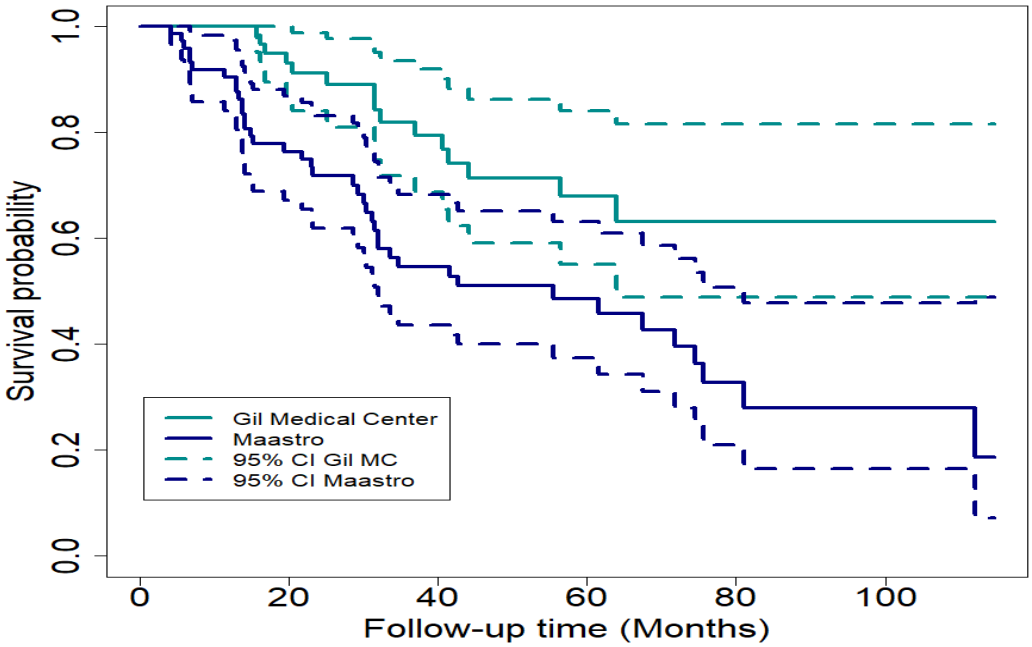

Figure 7.1. Kaplan–Meier OS curve for late-stage (IIB-IVA) cervical cancer patients with squamous cell carcinomas at Gil Medical Center(N = 70) and Maastro(N = 76).

Survival tree for OS

The decision tree was developed using all the essential independent prognostic factors for overall survival, as shown in Figure 7.2. The most important prognostic factor for determining overall survival was the patient's FIGO stage based on the tree. The tree then splits into patients with FIGO stage IIB tumor versus FIGO stage III-IVA tumor. The second most important factor for these two groups of patients was quite different. For stage IIB patients, age was the most significant factor, while for stage III-IVA patients, tumor size was more important. The survival curve for IIB patients above 70 years can be consulted directly from the tree. Likewise, stage III-IVA patients with a smaller than 4cm tumor. However, for IIB patients younger than 70 years and the more advanced stage patients (III-IVA) with a tumor larger than 4cm, the patient's SCC-Ag level has to be consulted but at slightly different concentrations of 3.75 *ng/ml* and 3.9 *ng/ml* respectively.

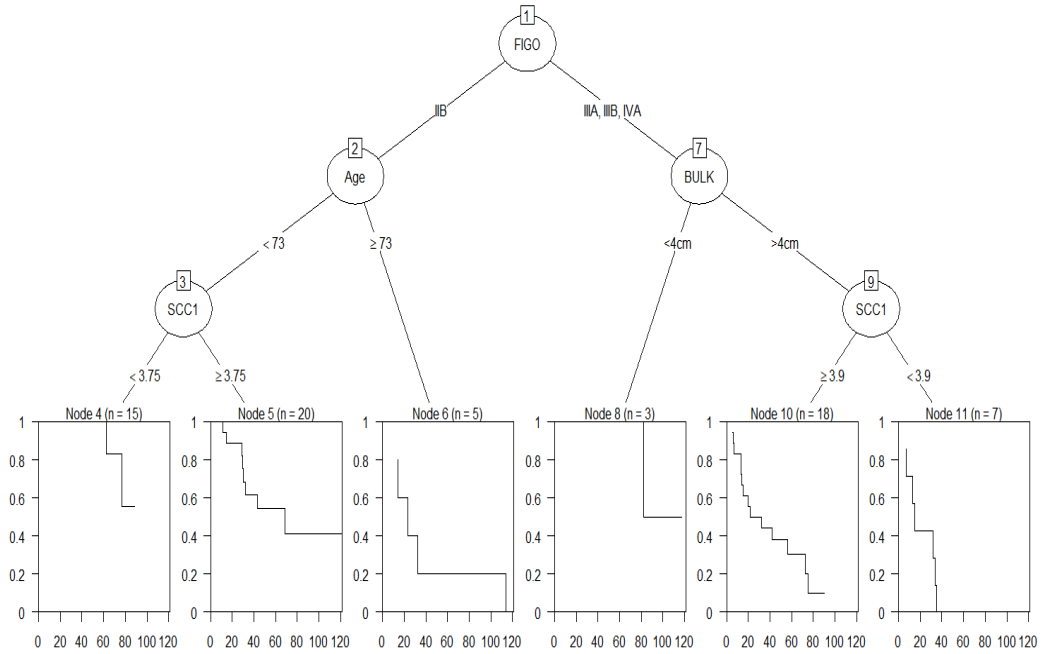

Figure 7.2. Survival tree for predicting the overall survival in women with stage IIB and higher cervical cancer.

The grown tree performs reasonably well in the training and validation data, with a C-index of 0.786 (95% CI, 0.711 - 0.860) and 0.716 (95% CI, 0.639 - 0.893). Six risk groups were created based on the splitting of the survival tree, as shown on the leaf nodes of Figure 7.2. The first three-leaf nodes of the tree were considered IIB patients with low, moderate, and high risk, and the subsequent three-leaf nodes were also considered III-VA patients with low, moderate, and high risk. The number of patients in each risk group is shown at the top of each leaf node's rectangular box. A Kaplan-Meier curve for each risk group is shown in Figure 7.3 for the training and validation cohort, respectively. The log-rank test for the difference between risk groups' survival curves was statistically significant in both cohorts, indicating that the survival tree is good at discriminating between patients based on their survival probabilities.

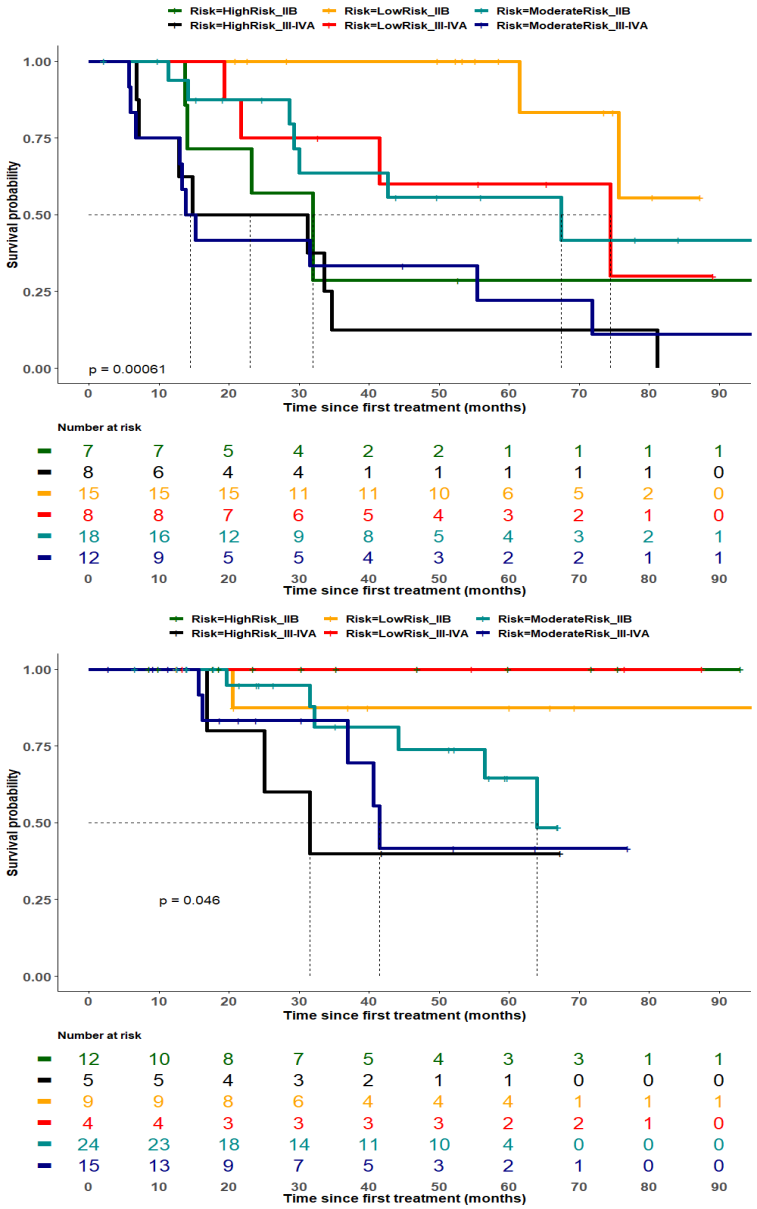

Figure 7.3. *The Kaplan-Meier survival curve for the different risk groups in the train and validation cohort*

Risk groups

To evaluate the decision tree's ability to predict future patients' survival probability in these different risk groups, the predicted mean survival curve of each risk group was compared with the observed Kaplan-Meier survival curves by overlaying the two plots, see Figure 7.4.

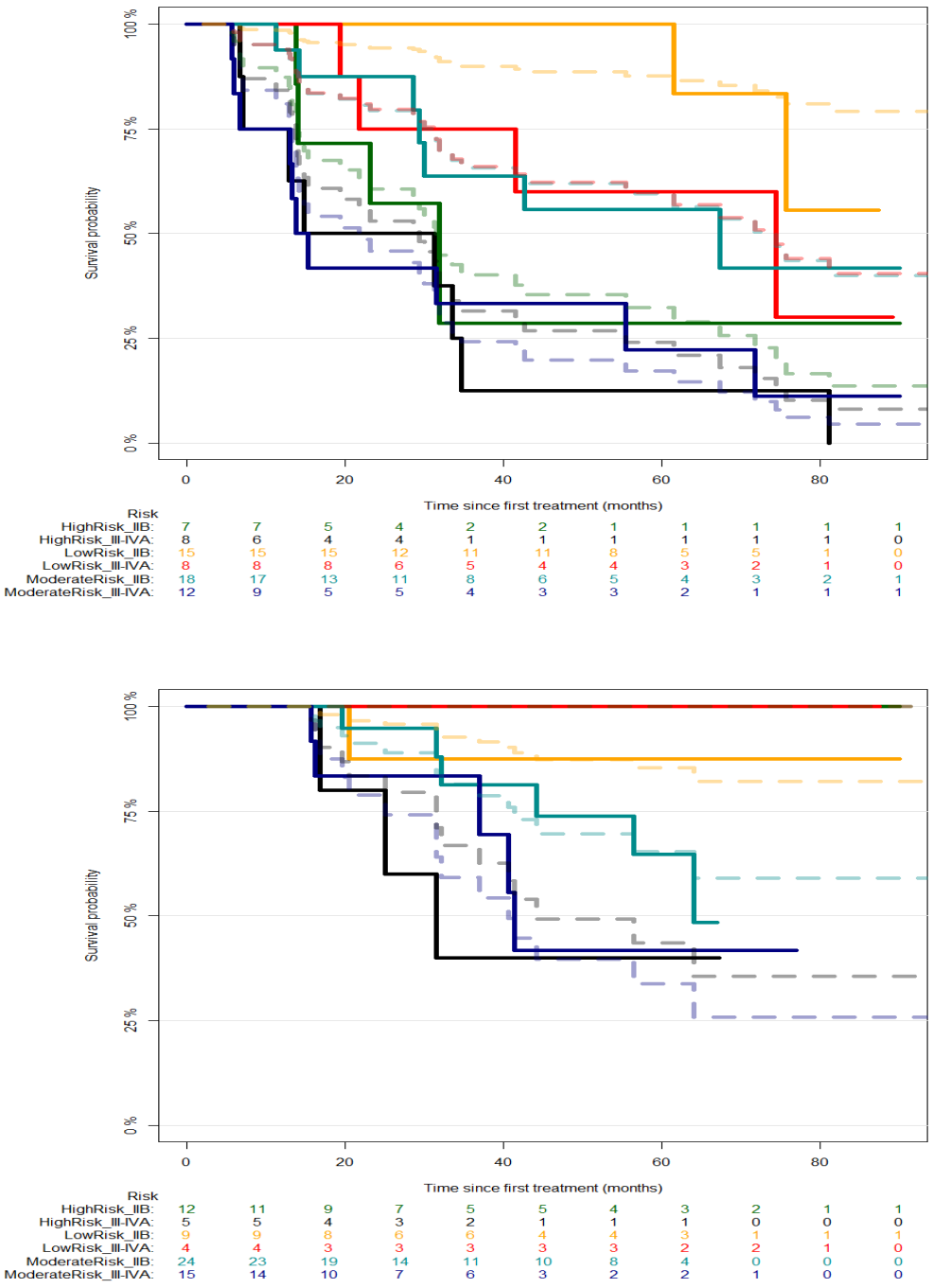

Figure 7.4. *Overlaid comparison of predicted mean survival curves (dotted lines) and observed Kaplan-Meier curve (solid lines) of the risk groups for the train and validation cohort respectively.*

Based on figure 7.4, the decision tree reasonably predicts these women's survival probability, given the close alignment of the actual and predicted survival curves in both the training and validation cohort.

Discussion

Survival is a significant concern and fear for most, if not all, persons diagnosed with cancer. In recent years, cervical cancer mortality has dropped significantly, with women at the very early stages of the disease having approximately a 100% survival rate (18). Although cervical cancer is treatable and even curable, it remains very lethal for some women, especially women at the advanced stage of the disease or in areas with limited medical care. Predicting the survival probability for these women is pivotal, given that most clinicians decide the therapeutic modality based on the patient's treatment outcome, especially overall survival. Also, some patients wish to know how long they have to live.

The developed survival tree performs reasonably well in the training and validation cohorts, with a C-index of 0.786 (95% CI, 0.711 - 0.860) and 0.716 (95% CI, 0.639 - 0.893). The survival tree's visual and straightforward graphical display with only four readily available clinical parameters (FIGO stage, age, tumor size, and SCC-Ag) means they can be readily printed on paper for clinical use. Also, the visual representation of the survival curve of patients in the different risk groups on the tree can be beneficial to physicians in their (shared) decision-making process.

Lymph node involvement was not significant enough for the tree to include it in predicting survival for these cervical cancer patients with squamous cell carcinomas. However, studies have shown that this variable is predictive of survival (19, 20). Yuan et al. (21), on the other hand, found in a study of approximately 800 cervical cancer patients that the tumor marker (SCC-Ag level) is significantly correlated with pelvic lymph node metastasis (PLNM) with a p-value of 0.001. The small sample size of this study and the significant correlation between lymph nodes and the SCC-Ag level might explain its absence on the survival tree. Generally, elevated pretreatment SCC-Ag level is associated with bulky tumors, treatment resistance, and poor survival (22–24). Therefore, this tumor-associated protein (SCC-Ag) is an all-important variable for survival prediction and detection of patients with possible tumor relapse at a very early stage, guiding physicians to select patients who might benefit most from adjuvant therapy.

Patient age was another significant survival predictor on the tree. However, the controversy surrounding the impact of age on the survival of cervical cancer women is overwhelming. Some studies suggest that younger age is an unfavorable prognostic factor (25, 26). In contrast, others believe that older women have an unfavorable prognosis (27), more so in the disease's advanced stages. On the other hand, Wright et al. (28), believe that age is a poor prognostic factor for cervical cancer patients. Despite literature suggesting that younger women may have a relatively improved outcome (29, 30), given the direct correlation of age with survival in various cancers (31, 32), with older people experiencing worse survival because of their comparatively weaker immune system.

The contradicting age results observed between these studies might be due to the patient population's heterogeneity enrolled or the demographic effect of these different studies.

For example, in this study, the significant age difference concerning survival seen between FIGO stage IIB patients at Maastro was absent for the other FIGO stages (IIIA-IVA) and utterly absent at the Gil medical center (Figure S7.1 supplemental material). However, the developed tree could still capture this heterogeneity within the data and present the knowledge graphically. As depicted on the tree, age is an essential predictive factor for survival but only for women with FIGO stage IIB, and older women above 70 have a relatively unfavorable survival.

We also evaluated the relationship between the continuous variables on the tree in both institutions. In general, the correlations between age and SCC-Ag level in both centers was not significant with more or less similar distribution pattern (Figure S7.2 supplemental material). However, age was positively correlated to the SCC-Ag level at Maastro and negatively correlated at the Gil medical center. A detailed plot of the correlation between age and SCC-Ag level stratified by FIGO stage (Figure S7.3 supplemental material) shows a significant difference between the two institutions, with FIGO stage IIIA and IVA having an opposite correlational relationship. This institutional difference based on the FIGO staging could also be seen in the boxplot of age with FIGO stage stratified by patients' survival status (Figure S7.1 supplemental material). Despite the considerable difference in the distribution of these variables between the two treatment institutions, the tree could still reasonably predict patients into the different risk groups.

Tree algorithms have many advantages for overcoming data problems and complexity (missing information, variable distribution, etc.) to provide caregivers with a visual tool to help them make data-driven decisions. However, most studies analyzing cervical cancer prognosis are not performed with the tree algorithm (23, 33–35). To the best of our knowledge, there has been no study to assess the predictive value of a decision tree in the survival of stage IIB-IVA cervical cancer patients with squamous cell carcinomas receiving CCRT. Importantly, this study's secondary value was to show the predictive value of decision trees and highlight their possible clinical use because caregivers who can benefit from them are less exposed or have very little knowledge of their existence and potential.

This study provides some exciting findings, including a reasonably good predictive performance. However, it is not void of limitations mainly because of its retrospective nature in which data quality is not optimal. Nevertheless, a detailed data quality check was performed by physicians on both data sets from the two institutions throughout the study, and patients with missing information were excluded rather than imputing missing data.

The sample size and the number of independent variables used for this study were quite limited, with only women included with the disease at an advanced stage. This means that the tree cannot be generalized to women in the early stage of the disease. However, most treatment institutions generally collect the variables used to grow the decision tree, and they have been established to be predictive for survival. The tree also performs reasonably well when validated on patients from another treatment institution where patients' characteristics differ significantly from the training population.

The tumor size was dichotomized to patients with tumors larger or smaller than 4cm, leading to information loss. However, tree algorithms work by binarizing the independent variables; hence, we could not salvage any information loss in this regard.

Conclusion

We have developed and externally validated a novel survival tree to predict the overall survival of locally advanced cervical cancer patients with squamous cell carcinomas and highlighted some essential values of the tree methodology for clinical application. The visual representation of the grown tree provides valuable survival insights for patients in the different risk groups, which healthcare providers can leverage for better patient management, given its predictive performance on both the train and validation data. However, we recommend an external validation of the tree on a larger sample size to ascertain its clinical applicability.

Supplemental materials

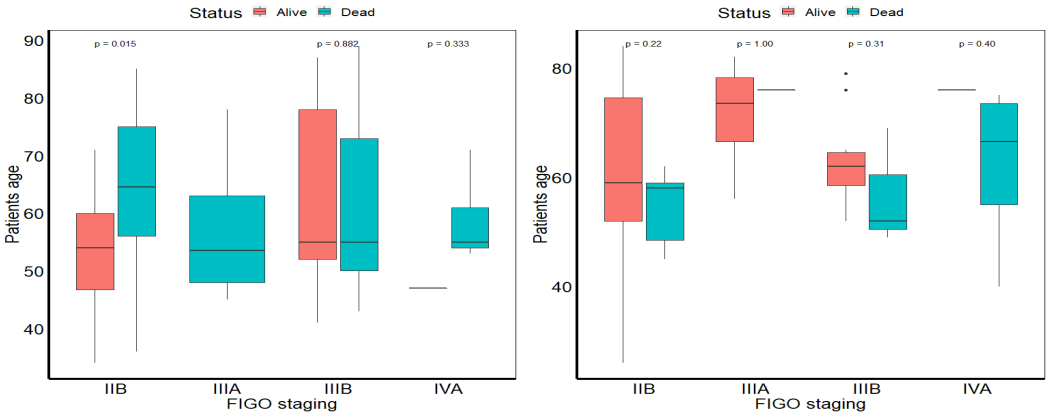

Figure 7.1. *Boxplot of age and FIGO stage by survival status for Maastro and Gil MC, respectively.*

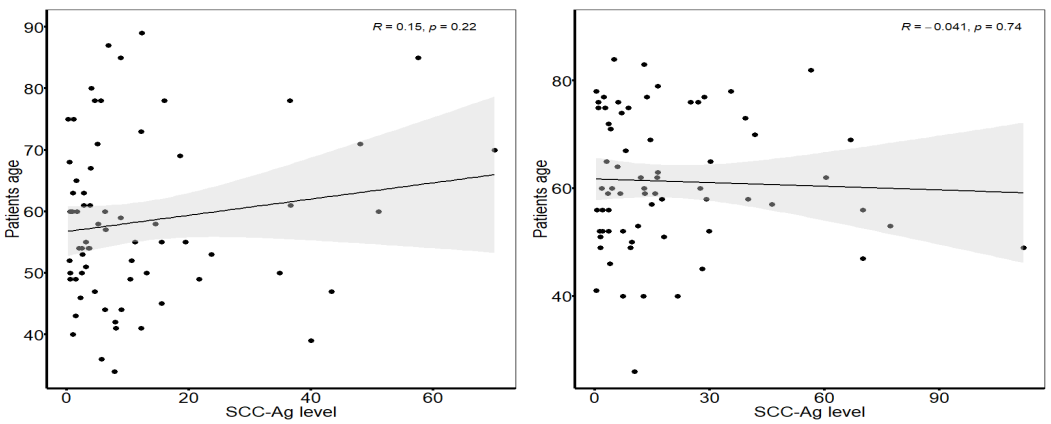

Figure 7.2. *Correlation plot of age and SCC-Ag level for Maastro and Gil MC, respectively.*

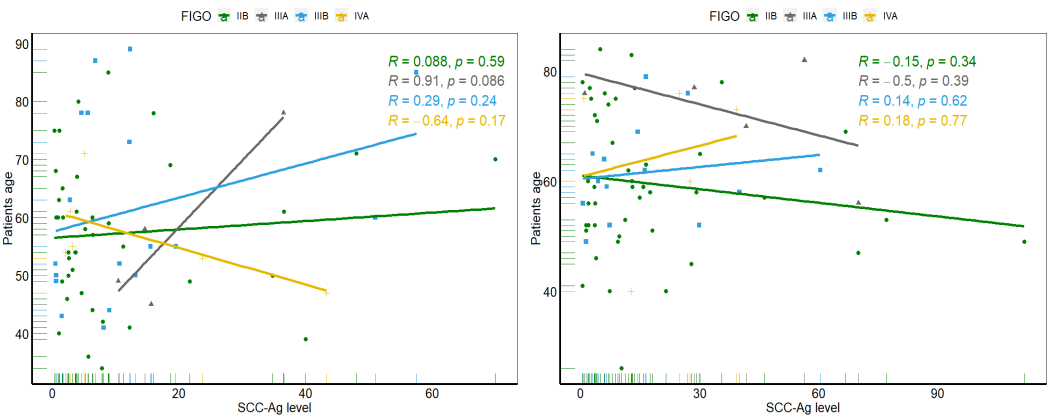

Figure 7.3. *Correlation plot of age and SCC-Ag by FIGO stage for Maastro and Gil MC, respectively.*

Bibliography

[1] Cezary Wojtyla, Kinga Janik-Koncewicz, and Carlo La Vecchia. Cervical cancer mortality in young adult european women. *European Journal of Cancer*, 126:56–64, 2020.

[2] National Cervical Cancer Coalition. cervical cancer overview, 2013. URL http://www.nccc-online.org/hpvcervical-cancer/cervical-cancer-overview/. Accessed: 2021-07-27.

[3] World Health Organization et al. Who guidance note: comprehensive cervical cancer prevention and control: a healthier future for girls and women. 2013.

[4] L Bruni, L Barrionuevo-Rosas, G Albero, B Serrano, M Mena, D Gómez, J Muñoz, FX Bosch, and S de Sanjosé. Ico/iarc information centre on hpv and cancer (hpv information centre). *Human Papillomavirus and Related Diseases in Netherlands. Summary Report 27 July 2017*, 27, 2017.

[5] Marjolein van Ballegooijen and RPMG Hermens. Cervical cancer screening in the netherlands, 2000.

[6] Human papillomavirus and related cancers, fact sheet 2018. URL https://hpvcentre.net/statistics/reports/NLD_FS.pdf?t=1594382872892. Accessed: 2021-10-22.

[7] Kyu-Won Jung, Young-Joo Won, Hyun-Joo Kong, and Eun Sook Lee. Prediction of cancer incidence and mortality in korea, 2019. *Cancer research and treatment: official journal of Korean Cancer Association*, 51(2):431, 2019.

[8] Carolyn Nakisige, Melissa Schwartz, and Anthony Okoth Ndira. Cervical cancer screening and treatment in uganda. *Gynecologic oncology reports*, 20:37–40, 2017.

[9] Peter G Rose, James Java, Charles W Whitney, Frederick B Stehman, Rachelle Lanciano, Gillian M Thomas, and Paul A DiSilvestro. Nomograms predicting progression-free survival, overall survival, and pelvic recurrence in locally advanced cervical cancer developed from an analysis of identifiable prognostic factors in patients from nrg oncology/gynecologic oncology group randomized trials of chemoradiotherapy. *Journal of Clinical Oncology*, 33(19):2136, 2015.

[10] S Polterauer, C Grimm, G Hofstetter, N Concin, C Natter, A Sturdza, R Pötter, C Marth, A Reinthaller, and G Heinze. Nomogram prediction for overall survival of patients diagnosed with cervical cancer. *British journal of cancer*, 107(6):918, 2012.

[11] Tae Hun Kim, Moon-Hong Kim, Beob-Jong Kim, Sang-Il Park, Sang-Young Ryu, and Chul-Koo Cho. Prognostic importance of the site of recurrence in patients with metastatic recurrent cervical cancer. *International Journal of Radiation Oncology* Biology* Physics*, 98(5):1124–1131, 2017.

[12] Marinela Capanu and Mithat Gonen. Building a nomogram for survey-weighted cox models using r. 2011.

[13] Shahrokh F Shariat, Umberto Capitanio, Claudio Jeldres, and Pierre I Karakiewicz. Can nomograms be superior to other prediction tools? *BJU international*, 103(4):492–497, 2009.

[14] Chen Hu and Jon Arni Steingrimsson. Personalized risk prediction in clinical oncology research: applications and practical issues using survival trees and random forests. *Journal of biopharmaceutical statistics*, 28(2):333–349, 2018.

[15] HP Lee, CC Lee, and YC Su. Decision tree and prediction model for hypopharyngeal cancer survival. *B-ENT*, 15(2):93–100, 2019.

[16] Harald Steck, Balaji Krishnapuram, Cary Dehing-oberije, Philippe Lambin, and Vikas C Raykar. On ranking in survival analysis: Bounds on the concordance index. In *Advances in neural information processing systems*, pages 1209–1216, 2008.

[17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

[18] Peng Chua et al. 100% 5-year survival rate in laparoscopic radical hysterectomy for early-stage cervical cancer is an achievable task. *Gynecology and Minimally Invasive Therapy*, 9(2):53–53, 2020.

[19] Hang Zhou, Xiong Li, Yuan Zhang, Yao Jia, Ting Hu, Ru Yang, Ke-Cheng Huang, Zhi-Lan Chen, Shao-Shuai Wang, Fang-Xu Tang, et al. Establishing a nomogram for stage ia-iib cervical cancer patients after complete resection. *Asian Pacific Journal of Cancer Prevention*, 16(9):3773–3777, 2015.

[20] Kyra N McComas, Anna M Torgeson, Bryan J Ager, Christopher Hellekson, Lindsay M Burt, Kathryn A Maurer, Teresa L Werner, and David K Gaffney. The variable impact of positive lymph nodes in cervical cancer: Implications of the new figo staging system. *Gynecologic oncology*, 156(1):85–92, 2020.

[21] CC Yuan, PH Wang, HT Ng, LC Tsai, CM Juang, and LM Chiu. Both tpa and scc-ag levels are prognostic even in high-risk stage ib-iia cervical carcinoma as determined by a stratification analysis. *European journal of gynaecological oncology*, 23(1):17–20, 2002.

[22] Zhenhua Liu and Hongtai Shi. Prognostic role of squamous cell carcinoma antigen in cervical cancer: a meta-analysis. *Disease Markers*, 2019, 2019.

[23] Tu Binbin, Wu Lingying, Huang Manni, An Jusheng, and Li Ning. Outcomes and prognostic factors of advanced squamous cervical cancer after concurrent chemoradiotherapy. *Zhonghua fu chan ke za zhi*, 49(5):348, 2014.

[24] Jingxuan Fu, Weiping Wang, Yidan Wang, Chengeng Liu, and Peichang Wang. The role of squamous cell carcinoma antigen (scc ag) in outcome prediction after concurrent chemoradiotherapy and treatment decisions for patients with cervical cancer. *Radiation Oncology*, 14(1):1–5, 2019.

[25] Myung-Hwa Jeong, Hyungjoo Kim, Tae-Hun Kim, Moon-Hong Kim, Beob-Jong Kim, and Sang-Young Ryu. Prognostic significance of pretreatment lymphocyte percentage and age at diagnosis in patients with locally advanced cervical cancer treated with definite radiotherapy. *Obstetrics & Gynecology Science*, 62(1):35–45, 2019.

[26] H-J Huang, T-C Chang, J-H Hong, C-J Tseng, H-H Chou, K-G Huang, and C-H Lai. Prognostic value of age and histologic type in neoadjuvant chemotherapy plus radical surgery for bulky ($\geq$ 4 cm) stage ib and iia cervical carcinoma. *International Journal of Gynecologic Cancer*, 13(2), 2003.

[27] Bridget A Quinn, Xiaoyan Deng, Adrianne Colton, Dipankar Bandyopadhyay, Jori S Carter, and Emma C Fields. Increasing age predicts poor cervical cancer prognosis with subsequent effect on treatment and overall survival. *Brachytherapy*, 18(1):29–37, 2019.

[28] Jason D Wright, Randall K Gibb, Sajeena Geevarghese, Matthew A Powell, Thomas J Herzog, David G Mutch, Perry W Grigsby, Feng Gao, Kathryn M Trinkaus, and Janet S Rader. Cervical carcinoma in the elderly: an analysis of patterns of care and outcome. *Cancer*, 103(1):85–91, 2005.

[29] Ruey-Jien Chen, Yu-Hung Lin, Chi-An Chen, Su-Cheng Huang, Song-Nan Chow, and Chang-Yao Hsieh. Influence of histologic type and age on survival rates for invasive cervical carcinoma in taiwan. *Gynecologic oncology*, 73(2):184–190, 1999.

[30] Yolanda van Der Graaf, Petronella GM Peer, Gerhard A Zielhuis, and Peter G Vooijs. Cervical cancer survival in nijmegen region, the netherlands, 1970–1985. *Gynecologic oncology*, 30(1):51–56, 1988.

[31] Pauline Wimberger, Nils Lehmann, Rainer Kimmig, Alexander Burges, Werner Meier, Berit Hoppenau, Andreas du Bois, et al. Impact of age on outcome in patients with advanced ovarian cancer treated within a prospectively randomized phase iii study of the arbeitsgemeinschaft gynaekologische onkologie ovarian cancer study group (ago-ovar). *Gynecologic oncology*, 100(2):300–307, 2006.

[32] Akiko Ioka, Hideaki Tsukuma, Wakiko Ajiki, and Akira Oshima. Influence of age on cervical cancer survival in japan. *Japanese journal of clinical oncology*, 35(8):464–469, 2005.

[33] T Basta, I Gawron, K Mirocki, D Babczyk, and R Jach. Analysis of histological type and staging of cervical cancer as prognostic factors among women treated in the department of gynecology and oncology, jagiellonian university in the years 2001-2014. *Przeglad lekarski*, 74(1):13–20, 2017.

[34] Daisuke Endo, Yukiharu Todo, Kazuhira Okamoto, Shinichiro Minobe, Hidenori Kato, and Noriaki Nishiyama. Prognostic factors for patients with cervical cancer treated with concurrent chemoradiotherapy: a retrospective analysis in a japanese cohort. *Journal of gynecologic oncology*, 26(1):12–18, 2015.

[35] Jingxuan Fu, Weiping Wang, Yidan Wang, Chengeng Liu, and Peichang Wang. The role of squamous cell carcinoma antigen (scc ag) in outcome prediction after concurrent chemoradiotherapy and treatment decisions for patients with cervical cancer. *Radiation Oncology*, 14(1):1–5, 2019.

Part V
Bayesian Network

CHAPTER 8

BAYESIAN NETWORK STRUCTURE FOR PREDICTING LOCAL TUMOR RECURRENCE

Authors:
Biche Osong, *Carlotta Masciocchi, Andrea Damiani, Inigo Bermejo, Elisa Meldolesi, Giuditta Chiloiro, Maaike Berbee, Seok Ho Lee, Andre Dekker, Vincenzo Valentini, Thierry Gorlia, Jean-Pierre Gerard, Claus Rödel, Krzysztof Bujko, Cornelis van de Velde, Joakim Folkesson, Aldo Sainato, Robert Glynne-Jones, Sam Ngan, Morten Brændengen, David Sebag-Montefiore, Johan van Soest.*

Adapted from: *Physics & Imaging in Radiation Oncology (**phiRO**)*
DOI: ***10.1016/j.phro.2022.03.002***

Abstract

Purpose
To develop an expert-elicited Bayesian network (BN) for predicting local tumor recurrence in rectal cancer and compare its performance with an algorithmically determined structure.

Summary of background
Tumor recurrence after treatment, a characteristic of malignant tumors, is the biggest concern for rectal cancer survivors. The epidemiology of the disease calls for a pressing need to improve healthcare quality and patient outcomes. Prediction models such as Bayesian networks (BN), which can probabilistically reason under uncertainty, could assist caregivers with patient management.

Patients and Methods
A retrospective study on 6754 patients with locally advanced rectal cancer (LARC) enrolled in 14 international clinical trials from 1993 to 2014. Local tumor recurrence at 2, 3, and 5 years was defined as the endpoints of interest. The expert BN structure was elicited from the opinions of five expert rectal cancer treating physicians from three countries on two continents. The algorithmic BN structure was developed from the data with the hill-climbing algorithm. The area under the curve (AUC) values and calibration plots were used to assess the performance of the structures.

Results
The median age of patients in this study was 61 (22 - 90) years, with males twice the number of females. The mean AUC of the expert structure on the training and validation data was above 0.9 and 0.8, respectively, for all the time points of interest. However, the algorithmic networks with 14 variables and more than 32 connections had superior performance over the expert structure with ten variables and 19 relationships over all-time points of interest.

Conclusion
We have developed and internally validated a Bayesian networks structure from experts' opinions, which can predict the risk of a LARC patient developing a tumor recurrence at 2, 3, and 5 years. Our result shows that algorithmic-based structures are more performant and less interpretable than expert-based structures.

Introduction

The introduction of total mesorectal excision (TME) surgery and the use of neoadjuvant chemoradiation (nCRT) have reduced mortality and recurrence rate for rectal cancer patients, with an incidence of locoregional relapses after treatments of 4-8% (1–3). Despite the low incidence, tumor recurrence remains the predominant concern for most rectal cancer survivors considering the relatively poor quality of life involved (4, 5). Besides the treatment procedure, several other factors such as tumor site, size, ethnicity, genetics, etc. could influence the chance of tumor recurrence (5–7) and processing all these pieces of information to estimate the likelihood of a patient developing a tumor recurrence after treatment can be overwhelming, even for experts (8, 9). Predictive models such as Bayesian networks, which consider causal relationships between features, can, on the other hand, learn efficiently from large and heterogeneous volumes of available information and make inferences about future patients.

Bayesian networks are suitable for clinical applications because they can probabilistically reason under uncertainty with an intuitive clinical interpretation of the results (10–12). Generally, they can be specified by an expert in the domain of interest or inferred from available data via a learning algorithm (11, 13). However, these methods may be challenging in healthcare. An algorithm-based structure can include spurious relationships that are not plausible or have no clinical meaning (e.g., causally linking gender to age) due to correlations in the data and the impossibility to determine the direction of causality from data (14). On the other hand, a structure specified by an expert might be biased by the expert's prior knowledge and subjective domain experience.

One possible solution to this problem is to survey multiple experts' opinions. This study hypothesizes that eliciting multiple experts' opinions will give a reliable Bayesian network structure to predict local tumor recurrences at several time points (2, 3, and 5 years) in rectal cancer patients whose predictions closely approximate the ground truth. To test this hypothesis, we implemented a solution to examine experts' opinions on the causal relationships to predict tumor recurrences in locally advanced rectal cancer (LARC) patients.

Materials and methods

A retrospective cohort of 6,754 diagnosed LARC patients treated with neoadjuvant chemoradiation followed by surgery from 1993 to 2014 from 14 international trial cohorts was analyzed for this study (Table S8.1 supplemental material). All the trials have different treatment protocols, patient characteristics, and accrual start dates. Only non-metastatic rectal cancer patients treated with conventional preoperative radiotherapy were considered for this study. Patients with a surgical procedure different from anterior-resection or abdominoperineal resection, treated with adjuvant radiotherapy or incomplete radiotherapy treatment, were excluded due to their relatively low representation. Figure S8.1 shows the variables under investigation in this study based on a timeline (**T**) of clinical practice availability. Local tumor recurrence at 2, 3, and 5 years was considered the endpoints of interest, defined as detecting a tumor on the same sites it previously started after therapy.

Statistics

Data from the 14 trial cohorts were merged and split into training and validation sets by performing a random 80 - 20% split (stratified per cohort). The SMOTE algorithm (15) was used to address the class imbalance per response time point, and continuous variables were categorized based on literature and experts' suggestions. The circumferential resection margin (CRM) was dichotomized into positive if the tumor is $\leq$1mm from the circumferential margin and negative if $>$1mm (Table S8.2 supplemental material). Missing values were considered as a category (Unknown) for all variables. However, patients with missing information on their local recurrence status were excluded from further analyses. All analyses were conducted in R version 3.6.1 (16) using the bnlearn package (17) and GeNIe (Graphical Network Interface) application (18) was used for structural visualization. Model performance was assessed by generating calibration plots, model accuracy, and calculating the area under the curve (AUC) on training and validation sets for all time points of interest.

The median age of 5404 patients in the training and 1350 in the validation cohorts was 61 (22 - 90) and 61 (25 - 84) years respectively. Patients characteristics and treatment modalities are shown in table 8.1

Table 8.1. General patient characteristics on the training and validation datasets

Variable	Levels	Training	Validation	p-value
Age (years)	Mean (sd)	61.4 (9.6)	61.4 (10)	0.82
	Missing	92 (1.7%)	17 (1.3%)	
Gender	Male	3760 (69.6%)	929 (68.8%)	0.52
	Female	1630 (30.2%)	420 (31.1%)	
	Missing	14 (0.2 %)	01 (0.1%)	
Clinical T	cT 1	93 (1.7%)	30 (2.2%)	0.13
	cT 2	387 (7.2%)	117 (8.7%)	
	cT 3	4002 (74.1%)	987 (73.1%)	
	cT 4	370 (6.8%)	84 (6.2%)	
	Missing	552 (10.2%)	132 (9.8%)	
Clinical N	cN 0	1547 (28.6%)	367 (27.2%)	0.57
	cN 1	1707 (31.6%)	438 (32.4%)	
	cN 2	303 (5.6%)	78 (5.8%)	
	Missing	1847 (34.2%)	647 (34.6%)	
RT dose (Gy)	Mean (sd)	47.7 (3.6)	47.7 (3.5)	0.98
	Missing	1378 (22.5%)	347 (22.9%)	
Surgery procedure	APR	1629 (30.1%)	426 (31.6%)	0.42
	ARbased	3489 (64.6%)	851 (63.0%)	
	No surgery	107 (2.0%)	22 (1.6%)	
	Missing	179 (3.3%)	51 (3.8%)	
Circumferential resection margin	Negative	543 (10.1%)	140 (10.3%)	0.91
	Positive	435 (8.0%)	114 (8.4%)	
	Missing	4426 (81.9%)	1096 (81.2%)	
Overall treatment time (days)	Mean (sd)	37 (6.6)	37.4 (9.3)	0.16
	Missing	1598 (26.1%)	396 (25.8%)	
Neoadjuvant chemo	5FU+OXI	1128 (20.9%)	266 (19.7%)	0.60
	5FUbased	2806 (51.9%)	709 (52.5%)	
	No Chemo	1245 (23.0%)	321 (23.8%)	
	Missing	225 (4.2%)	54 (4.0%)	
Tumor distance[d](cm)	Mean (sd)	06 (3.1)	06 (3.1)	0.84
	Missing	1023 (16.7%)	260 (17.0%)	
Interval between RT and Surgery (weeks)	Mean (sd)	0.9 (0.4)	0.9 (0.3)	0.77
	Missing	2251 (36.7%)	554 (36.2%)	
Adjuvant Chemo	5FU+OXI	651 (12.0%)	152 (11.3%)	0.70
	5FUbased	2497 (46.2%)	621 (46.0%)	
	No Chemo	2024 (37.5%)	515 (38.1%)	
	Missing	232 (4.3%)	62 (4.6%)	
Pathological N	ypN 0	3436 (63.6%)	852 (63.1%)	0.95
	ypN 1	1225 (22.7%)	311 (23.0%)	
	ypN 2	312 (5.7%)	77 (5.7%)	
	Missing	431 (8.0%)	110 (8.2%)	
Pathological T	ypT 0	625 (11.5%)	148 (11.0%)	0.05
	ypT 1	307 (5.7 %)	95 (7.0%)	
	ypT 2	1453 (26.9%)	387 (28.7%)	
	ypT 3	2413 (44.7 %)	557 (41.3%)	
	ypT 4	175 (3.2 %)	53 (3.9%)	
	Missing	431 (8.0%)	110 (8.1%)	
2 years local recurrence	True	385 (7.1%)	90 (6.7%)	0.49
	False	4168 (77.1 %)	1060 (78.5%)	
	Missing	851 (15.8%)	200 (14.8%)	
3 years local recurrence	True	487 (9.0%)	118 (8.8%)	0.61
	False	3445 (63.8%)	882 (65.3%)	
	Missing	1472 (27.2%)	350 (25.9%)	
5 years local recurrence	True	599 (11.1%)	153 (11.3%)	0.66
	False	2036 (37.7%)	497 (36.8%)	
	Missing	2769 (51.2%)	700 (51.9%)	

sd: standard deviation, **d**: Distance to anal verge (cm), **Chemo**: Chemotherapy, **RT**: Radiotherapy
APR: Abdominoperineal resection, **ARbased**: Anterior resection, **OXI**: oxaliplatin, 5**FU**: $5-$Fluorouracil

Structure learning

The domain knowledge from multiple experts' in three international radiotherapy institutions (Gemelli, Maastro and Gil Medical Center) was employed to develop and validate the Bayesian network structure. Two experts from Gemelli independently defined the causal relationship between the variables. These experts were requested to draw arrows between variables to indicate causal relationships without setting the relationships' importance. The relationships were restricted to only variables in the same time-point **t** or the next **t + n**. Arrows drawn from a given variable at time point **t** to another variable in a preceding time point **t - n** were considered invalid. For example, arrows from **Clinical T stage** to **Clinical N stage** or from **Age** to **Clinical T stage** are accepted. However, arrows from **Clinical T stage** to **Age** are rejected.

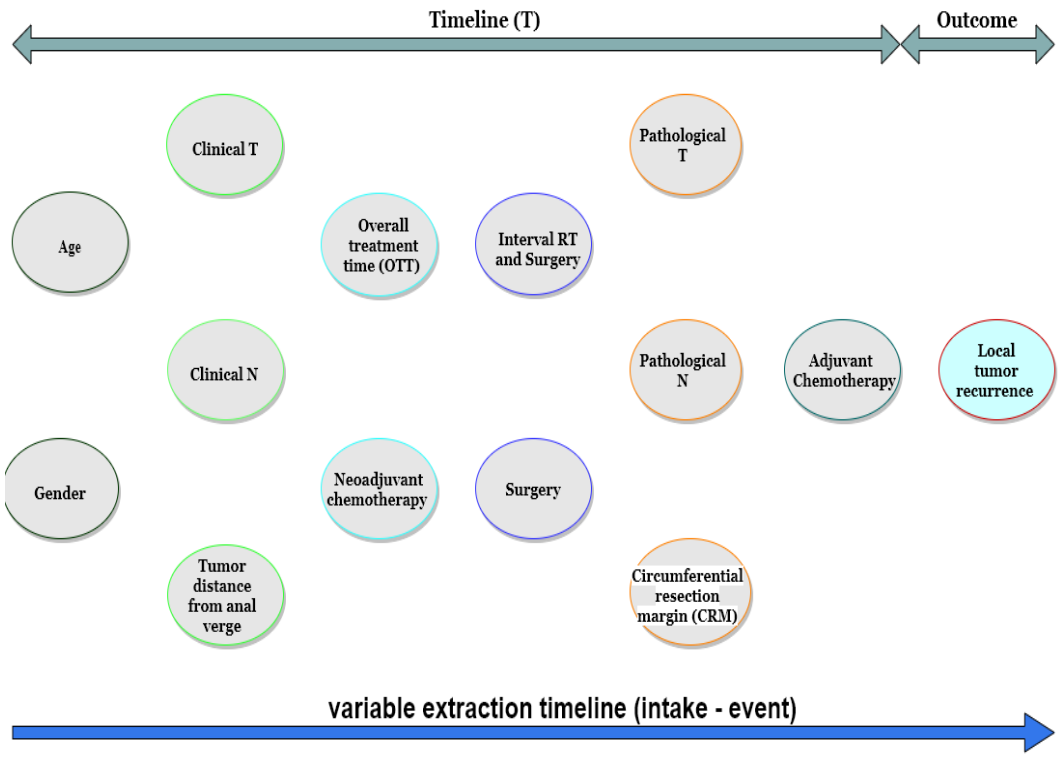

Figure 8.1. Variables under investigation on extraction timeline

Two other experts from Maastro separately reviewed the subset of connections common to both experts from Gemelli. The Dutch experts were tasked to validate the relationship by agreeing or disagreeing with each of the connections between the variables made by Gemelli experts. Only connections where both experts agreed were considered for further evaluation. As a final validation, an expert from Gil Medical Center reviewed the subset of connections common to both experts from Maastro. Only the connections validated by the expert were used to construct the final structure. The expert-developed structure was checked for cycles, which are not allowed in Bayesian network structures.

Structure comparison

In order to compare the performance of the developed expert Bayesian network structure in predicting local recurrence in rectal cancer patients, a structure was also inferred solely from the data with the hill climbing (HC) algorithm (19) for each time point of interest using the same training and validation data as the expert structure. The structures were first compared structurally and then numerically using the AUC, sensitivity, and specificity values. Calibration plots that measure how similar the distribution and behavior of the predicted probabilities are to that observed in data were produced to further evaluate the performance of the structure. The HC algorithm, which looks for the best structure over the search space by adding, removing, and reversing arcs (arrows) in the DAG one at a time, was preferred because it is computationally efficient, and a random restart search was implemented to prevent the structure from getting stuck on a bad local optimum (20). The Bayesian Information Criterion (BIC), a statistical goodness-of-fit measure that penalizes structural complexity, was used for the structure-learning process (19).

Results

Figure 8.2 shows the resulting BN structure based on expert knowledge. This network achieved AUCs above 0.9 and 0.8 for predicting the risk of local recurrence on the training and validation data, respectively, for all time points of interest. Table 8.2 shows the mean accuracies, AUCs, and confidence intervals of the structure's performance on the training and validation data at all follow-up time-points of interest.

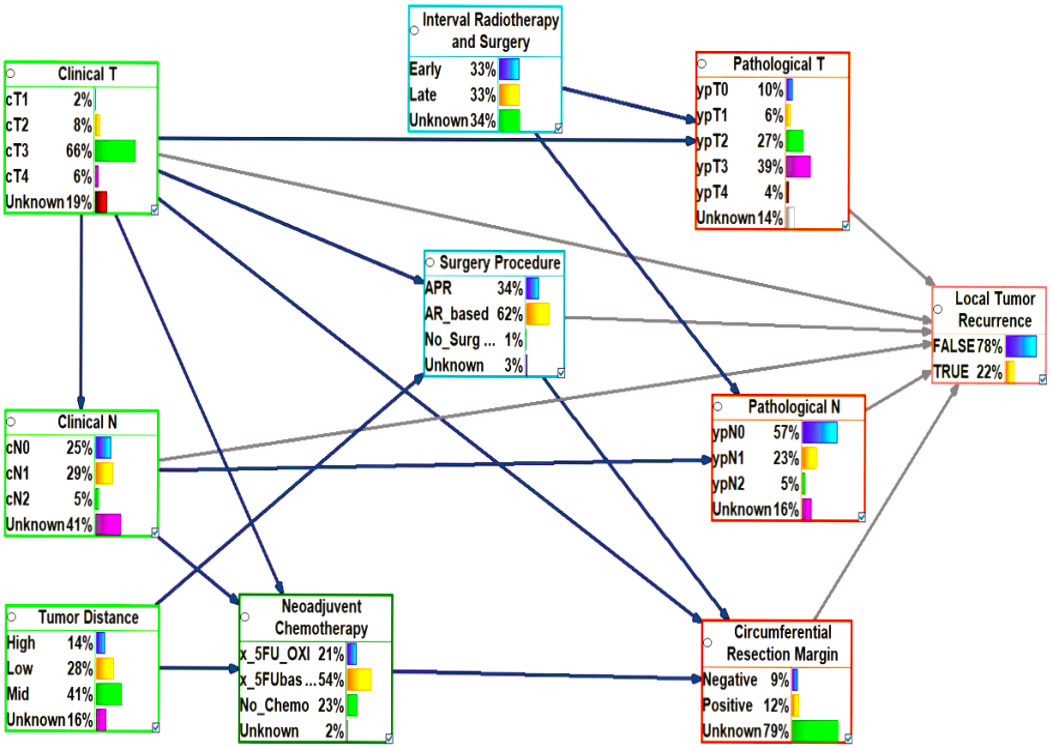

Figure 8.2. *Bayesian network structure based on expert knowledge. The boxes represent the variables (Node); the colors represent the variables' time points (t) of availability in the clinical process, as shown in Figure 8.1. The arrows indicate cause-effect relationships. The gray arrows indicate a direct causal effect on the outcome of interest.*

Table 8.2. The performance of the expert structure based on the accuracy and AUC values at different time points on the training and validation data.

	Training			Validation		
Time	ACC	AUC	95% CI	ACC	AUC	95% CI
2 years	0.84	0.92	0.91 - 0.93	0.76	0.87	0.86 - 0.88
3 years	0.83	0.92	0.91 - 0.92	0.74	0.86	0.84 - 0.87
5 years	0.84	0.91	0.92 - 0.93	0.72	0.80	0.79 - 0.82

CI: confidence interval, **ACC:** Accuracy

Structure comparison

The Bayesian network structures resulting from the HC algorithm mentioned above used all 14 available variables (Figure S8.1 supplemental material) with 32, 33, and 30 arcs for 2, 3, and 5 years tumor recurrence, respectively. On the other hand, the experts' structure had 19 arcs and ten variables, excluding age, gender, adjuvant chemotherapy, and overall treatment time. The outcome had a direct patent-to-child connection with all 13 nodes for the structure at 5-years, and the outcome for the 2-year structure had 11 children, excluding the arc with adjuvant chemotherapy and CRM, while that of the 3-year structure had 10, excluding adjuvant chemotherapy, pathological T, and N which was quite the opposite for the expert structure with just six parents. The only similarity between the algorithmic and expert structures was the arc CRM to the outcome for the 2-year structure and the arcs pathological T and N to the outcome for the 3-year structure.

Based on the relationship of the variables with the outcome among the algorithmic structures, the arcs pathological T and N to the outcome for the 3-year structure were revered in the 2, and 5-year structures and the arc CRM to the outcome in the 2-year structure was reversed in the 3 and 5-year structures while the arc adjuvant chemotherapy to the outcome in the 2-year structure was revered in the 5-year structure and absent in the 3-year structure. Age was only connected to the outcome alone for 3 and 5 years structures.

The numerical comparison of the expert and algorithmic structures based on their performance in terms of the AUC (Figure S8.2 supplemental material), sensitivity, and specificity values for each time points in the training and validation data is shown in Table 8.3. The algorithmic structures performed slightly better than the expert structure for all matrices of interest, especially in the validation data. However, these differences in the AUC values between the structures were statistically significant (p-value < 0.05) for all time points.

Table 8.3. The AUC, sensitivity, and specificity values of the expert and algorithmic structures on the training and validation data at different time points.

Training				
Time	**Structure**	**AUC**	**Sensitivity**	**Specificity**
2 years	Experts	0.92	0.93	0.76
	Algorithm	0.93	0.92	0.78
3 years	Experts	0.92	0.91	0.75
	Algorithm	0.94	0.92	0.80
5 years	Experts	0.92	0.92	0.76
	Algorithm	0.94	0.92	0.80
Validation				
2 years	Experts	0.87	0.86	0.67
	Algorithm	0.90	0.88	0.75
3 years	Experts	0.86	0.84	0.63
	Algorithm	0.92	0.89	0.76
5 years	Experts	0.80	0.73	0.71
	Algorithm	0.89	0.96	0.72

The calibration plots in Figure 8.3 show a good match between the predicted probabilities and the observed frequencies in the training and validation cohort for both expert and algorithmic structures. Generally, the expert structure seems to be better calibrated than the algorithmic structures since most of its points are closer to the dotted diagonal gray line representing an ideal model, especially recurrence at 5 years.

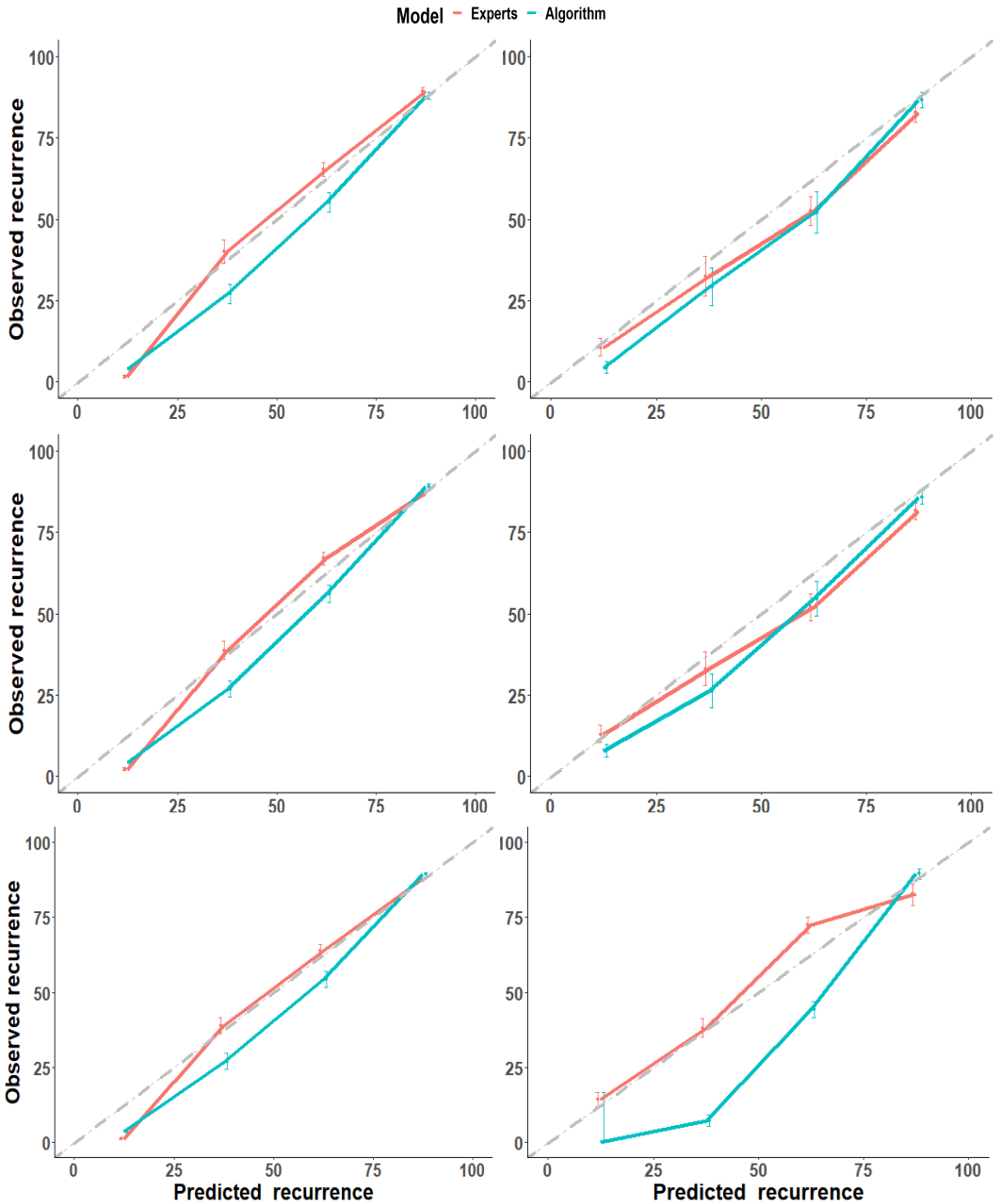

Figure 8.3. *Calibration plots of the models on the training (left) and validation (right) data for 2-year (top) to 5-year (bottom) local recurrence. The gray dashed line represents ideal calibration, while solid lines represent each model's calibration. Vertical bars indicate a 95% confidence interval, and dots indicate bias-corrected estimates.*

Discussion

In the current study, we have developed and internally validated a Bayesian network structure based on five experts' opinions from three different radiation therapy treatment institutions to predict tumor recurrence for locally advanced rectal cancer patients. The structure was developed to capture the biological process that leads to a tumor recurrence by connecting the variables in the structure based on a timeline of their clinical availability. The developed structure used nine clinical features well-known to clinicians as pivotal factors for predicting local tumor recurrence. The structure was well-calibrated with predictive performance based on the AUC values above 0.9 and 0.8 in training and validation data, respectively, for all time points of interest. Structures inferred from a learning algorithm used four variables more than the expert structure, including age, gender, adjuvant chemotherapy, and overall treatment time, with slightly better performance in terms of AUC values. However, the expert structure was clinically more plausible than the algorithmic structures and aligned with the clinical process.

The choice of model to predict local tumor recurrence in rectal cancer patients in this study was influenced by two main reasons. Firstly, Bayesian networks better represent complex systems such as the clinical processes leading to tumor recurrence since they have more liberty to define interactions between variables (10, 21), unlike the generally used regression method (22–24). Secondly, their ability to make inferences on any variable(s) in the network makes them very valuable for decision support as they can serve as a diagnostic and prognostic tool. Thus far, this study is the first to assess the predictive value of Bayesian networks for tumor recurrence in rectal cancer patients which makes a direct comparison and interpretation with other studies arduous because of the difference in analytical approach and study design.

Nonetheless Valentini et al. (22) have previously developed a model to predict tumor recurrence in locally advanced rectal cancer patients as a decision support tool. The model included age as a predictive factor for local recurrence, a variable missing in our expert structure. One apparent reason for this difference is the variable selection procedure, given that age is also present in the algorithmic structures. Algorithm-based variable selection methods exploit spurious correlations within the data, which might unnecessarily increase the complexity of the model and cause overfitting. Also, Farhat et al. (25) did not find age or gender a predictive factor for tumor recurrence in their 16-year respective study, which is in support of the expert structure. Although the controversy of included variables depends largely on patient heterogeneity between the studies, variable inclusion in a prediction model must not always be strictly algorithm-based, even if it improves its performance. Instead, it should contain some level of clinical understanding, context, or rationale, which domain experts are more suited to provide because a correlation between variables does not necessarily imply causality.

The algorithmic structures used four variables more than the expert structure, which explains the slightly better predictive performance. However, the additional variables increase the complexity of the structure since more connections between the variables are formed but with minimal predictive benefits (Table 8.3). Although the expert structure did not perform better than the algorithmic structures, we believe the expert structure might be more suitable for clinical use over the algorithmic structure. Firstly, the algorithmic structure uses parent-to-child connections not aligned with the clinical process (e.g., the pathological nodal stage has a causal link to age at start radiotherapy) to make decisions. This implies that even with better performance on both data sets, algorithmically generated structures are comparable to black-box models since the decision process lacks clinical explanation. Regarding model calibration, the expert structure seems to be better calibrated for all time points than the algorithmic structure with higher AUCs.

Tumor recurrence is a very challenging endpoint not only in terms of quality of life for cancer survivors (4, 5) but also the difficulty in accurately predicting the endpoint (26). Patient variability can explain this difficulty since a treatment regime that leads to recurrence-free for one patient might not give another patient the same outcome. Therefore, collaboration with domain experts is pivotal to have a more personalized prediction of tumor recurrence since they better understand tumor biology. The performance of the proposed expert structure is well above the chance level with clinically valid relationships. Therefore, it might be valuable in routine clinical settings as a decision aid to support personalized treatment decision-making. Also, it could guide clinicians to opt for a more aggressive adjuvant therapy to prevent the chance of a tumor recurrence for patients who have undergone surgery but with a high predicted probability of a tumor recurrence in the structure. However, the structure is trained on retrospective clinical trial data and warrants an external validation on routine clinical data to ascertain its clinical usefulness. In addition, the circumferential resection margin, a variable proven to influence local recurrence (27, 28), had a large proportion of missing information and will be worthwhile to retrain the CPT of this variable on more complete dataset.

Despite the predictive performance of the Bayesian network structures in this study, there is still room for improvement. The international and multi-trial nature of our study may be seen as a limitation, given that it combines the contribution of experts from three cancer institutions and data from multiple clinical trials with different treatment protocols. However, this limitation could also be considered a strength, as it may make our findings more robust and generalizable. The multi-trial combination is particularly relevant for this study since it enables the structure to be trained on a large sample size, which reflects the models' superior performance over other studies with relatively smaller sample sizes (22–24) given that model performance is proportional to training sample size. Also, this large sample size helped improve the number of events given the disease's low event

rate. Despite combining data from 14 different European trials, the number of local recurrence events was relatively low. Our study design also prevents updating the number of connections between variables since each expert is contacted only once for input. Also, some of the variables were categorized, leading to a loss of information. Lastly, blood tumor markers such as carcinoembryonic antigen (CEA), a proven predictive factor for local tumor recurrence in rectal cancer (25), were not included in the structure because of the study's retrospective nature.

Conclusions

we have developed and validated a Bayesian network structure from 14 trial cohorts' data by analyzing a total number of 6754 rectal cancer patients for predicting the risk of local recurrence in locally advanced rectal cancer patients at 2, 3, and 5 years. The causal relationships between the variables in the developed Bayesian network structure were proposed and validated by domain experts with years of experience from different international radiotherapy centers, where treatment protocols may differ. Our result showed that although structures from both methods performed above chance level, the algorithmic based structures had higher discriminating power than the expert structure. However, they contained clinically incomprehensible arcs, making the expert structure more credible even with relatively lower predictive performance. Future research will combine these two Bayesian network structure learning approaches to produce clinically plausible structures with optimal predictive performance.

Supplemental materials

Table 8.1. The 14 clinical trials, accrual dates and the total number of patients from each trial included in this study

Trial	Accrual	Total	Trial description
EORTC-22921-ROG	1993 - 2003	1006	Four arms phase III clinical trial for T3-T4 resectable rectal cancer comparing pre-operative pelvic irradiation to pre-operative irradiation combined with fluorouracil and Leucovorin with or without post-operative adjuvant chemotherapy
Polish I	1999 - 2002	139	Long-term results of a randomized trial comparing preoperative short-course radiotherapy with preoperative conventionally fractionated chemo-radiation for rectal cancer
ACCORD	2005 - 2008	594	The ACCORD 12 trial investigated the value of two different preoperative chemo-radiotherapy (CT-RT) regimens in T3-4 Nx M0 resectable rectal cancer.
INTERACT	2006 - 2013	476	Long-term results of a randomised trial on preoperative capecitabine based radiochemotherapy intensified by concomitant boost or oxaliplatin, for cT2 (distal)–cT3 rectal cancer
MRC CR07	1998 - 2005	717	Compared short-course preoperative radiotherapy versus initial surgery with selective postoperative chemoradiotherapy.
FFCD 2903	1993 - 2003	732	Preoperative Radiotherapy With or Without Concurrent Fluorouracil and Leucovorin in T3-4 Rectal Cancers
CAO/ARO/AIO-94	1995 - 2002	503	Compared preoperative chemoradiotherapy with postoperative chemoradiotherapy for locally advanced rectal cancer
Polish II	2008 - 2014	267	Long-course oxaliplatin-based preoperative chemoradiation versus 5 $\times$ 5 Gy and consolidation chemotherapy for cT4 or fixed cT3 rectal cancer
I-CNR-RT	1992 - 2001	605	Evaluate the effect of adjuvant chemotherapy (ACT) in locally advanced rectal cancer (LARC) after neoadjuvant chemoradiation (NACT-RT).
Nordic	1996 - 2003	205	Compared preoperative long-course radiotherapy alone (RT) or with chemotherapy (CRT) in the most locally advanced/ugly rectal cancers.
Chronicle	2005 - 2008	111	Examined the benefit of postoperative adjuvant capecitabine and oxaliplatin (XELOX) chemotherapy
CAO/ARO/AIO-04	2008 - 2010	1222	Preoperative chemoradiotherapy and postoperative chemotherapy with fluorouracil and oxaliplatin versus fluorouracil alone in locally advanced rectal cancer
Dutch	1996 - 1999	14	To document local recurrence in primary rectal cancer when standardised techniques of surgery, radiotherapy, and pathology are used, and to investigate whether the local recurrence rate after total mesorectal excision permits the omission of adjuvant short term preoperative radiotherapy.
TROG 01-04	2001 - 2006	163	Randomized trial of short-course radiotherapy versus long-course chemoradiation comparing rates of local recurrence in patients with T3 rectal cancer

Table 8.2. Variables categorized into different groups

Variable	**Levels**	**Names**
Patient Age at Radiotherapy	Less than 49 years	Adults
	Between 50 to 59 years	Aged
	Between 60 to 69 years	Old
	Above 70 years	Senior
Interval between radiotherapy and surgery	Before six weeks	Early
	After six weeks	Late
Tumor distance from the anal verge in cm	Less than 5 cm	Low
	Between 5 - 10 cm	Mid
	Above 10 cm	High
Overall treatment time (OTT)	Less than 37 days	Short
	More than 37 days	Long
Circumferential resection margin (CRM)	$\leq$ 1mm	Positive
	$>$ 1mm	Negative

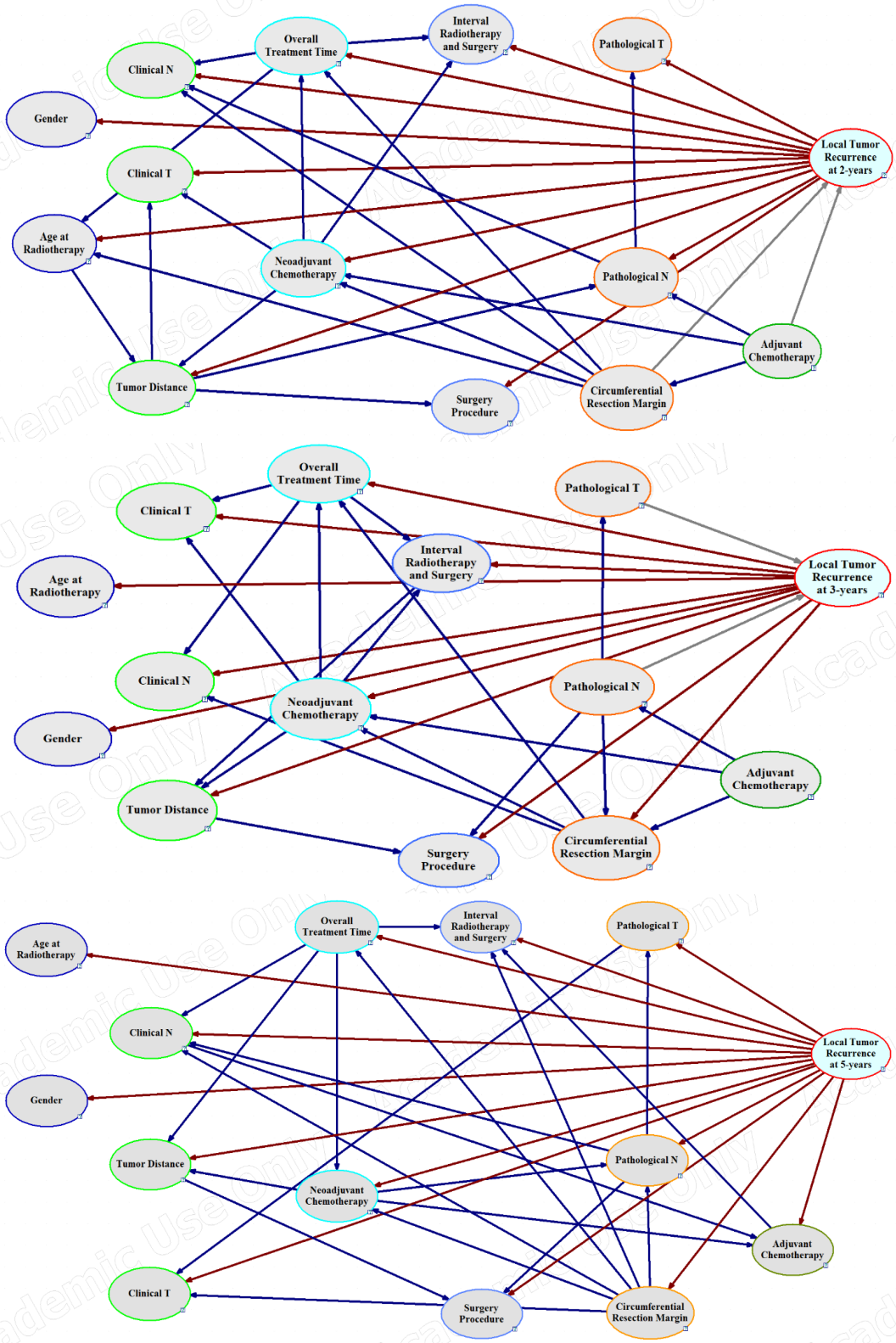

Figure 8.1. *The Bayesian network structures based on the hill-climbing algorithm. The circles represent the variables (Node), and the arrows indicate the direction of the causal-effect relationships. The brown arrows indicates a direct causal-effect from the outcome of interest*

Acknowledgments

The authors gratefully acknowledge the many individuals who contributed to the fruitfulness of this publication, with special thanks to the trial maintainers for their permission to use the data from their trials for this research. However, some of the results of this publication were presented at the 39th annual meeting of The European SocieTy for Radiotherapy and Oncology (ESTRO) online, from 28 November to 01 December 2020. The contents of this publication and methods used are solely the responsibility of the authors and do not necessarily represent the official views of the EORTC. We are also grateful to the anonymous reviewers of this manuscript for their valuable comments and suggestions that have greatly helped us to improve the paper.

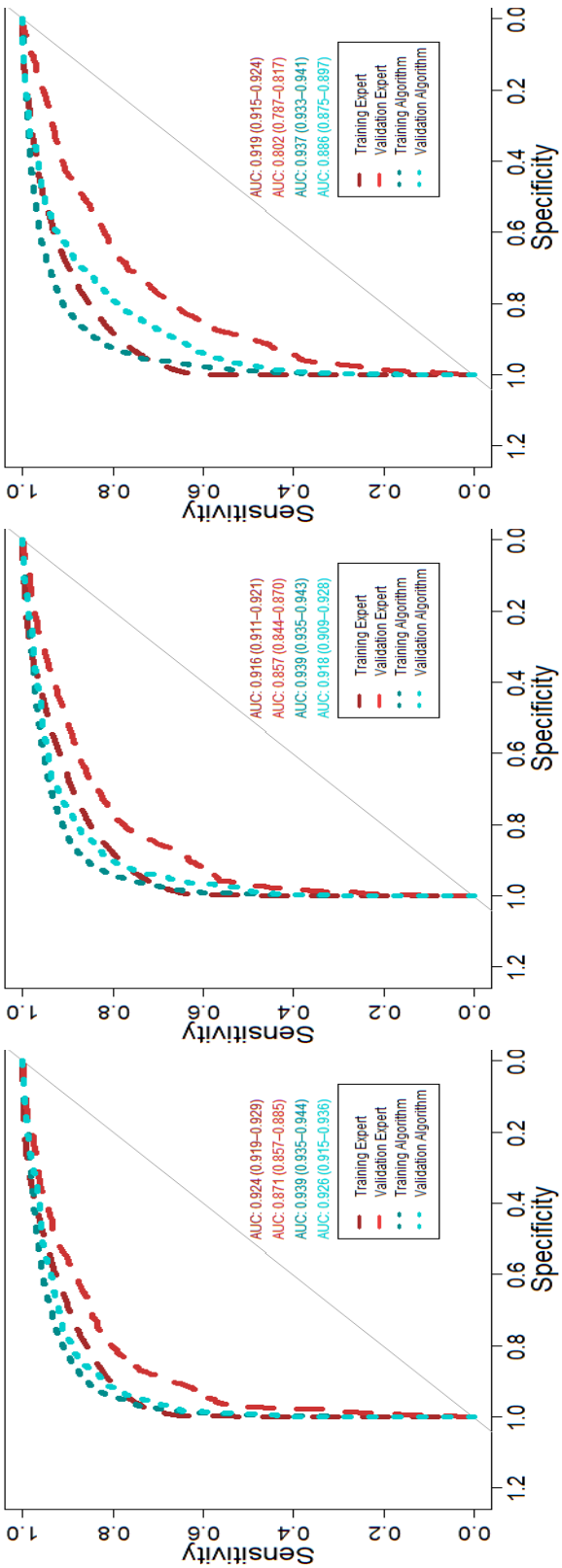

Figure 8.2. *ROC curves for expert (red) and algorithm (cyan) structures on the training and validation data for 2-years(top) to 5-years (bottom)*

Bibliography

[1] David Sebag-Montefiore, Richard J Stephens, Robert Steele, John Monson, Robert Grieve, Subhash Khanna, Phil Quirke, Jean Couture, Catherine de Metz, Arthur Sun Myint, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (mrc cr07 and ncic-ctg c016): a multicentre, randomised trial. *The Lancet*, 373:811–820, 2009. doi: https://doi.org/10.1016/S0140-6736(09)60484-0.

[2] Ellen Kapiteijn, Corrie AM Marijnen, Iris D Nagtegaal, Hein Putter, Willem H Steup, Theo Wiggers, Harm JT Rutten, Lars Pahlman, Bengt Glimelius, J Han JM van Krieken, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *New England Journal of Medicine*, 345:638–646, 2001. doi: 10.1056/NEJMoa010580.

[3] Naruhiko Ikoma, Y Nancy You, Brian K Bednarski, Miguel A Rodriguez-Bigas, Cathy Eng, Prajnan Das, Scott Kopetz, Craig Messick, John M Skibber, and George J Chang. Impact of recurrence and salvage surgery on survival after multidisciplinary treatment of rectal cancer. *Journal of Clinical Oncology*, 35:26–31, 2017. doi: 10.1200/JCO.2016.72.1464.

[4] Frank Baker, Maxine Denniston, Tenbroeck Smith, and Michele M West. Adult cancer survivors: how are they faring? *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 104:2565–2576, 2005. doi: 10.1002/cncr.21488.

[5] Crystal S Denlinger and Andrea M Barsevick. The challenges of colorectal cancer survivorship. *Journal of the National Comprehensive Cancer Network*, 7:883–894, 2009. doi: 10.6004/jnccn.2009.0058.

[6] Aylwyn Mannell. An overview of risk factors for recurrent breast cancer. *South African Journal of Surgery*, 55:29–34, 2017.

[7] Mohammad Zare-Bandamiri, Mohammad Fararouei, Shadi Zohourinia, Nima Daneshi, and Mostafa Dianatinasab. Risk factors predicting colorectal cancer recurrence following initial treatment: a 5-year cohort study. *Asian Pacific journal of cancer prevention: APJCP*, 18:2465, 2017. doi: 10.22034/APJCP.2017.18.9.2465.

[8] Amy P Abernethy, Lynn M Etheredge, Patricia A Ganz, Paul Wallace, Robert R German, Chalapathy Neti, Peter B Bach, and Sharon B Murphy. Rapid-learning system for cancer care. *Journal of Clinical Oncology*, 28:4268, 2010. doi: 10.1200/JCO.2010.28.5478.

[9] Cary Oberije, Georgi Nalbantov, Andre Dekker, Liesbeth Boersma, Jacques Borger, Bart Reymen, Angela van Baardwijk, Rinus Wanders, Dirk De Ruysscher, Ewout Steyerberg, et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. *Radiotherapy and Oncology*, 112:37–43, 2014. doi: 10.1016/j.radonc.2014.04.012.

[10] Kevin B Korb and Ann E Nicholson. *Bayesian artificial intelligence*. CRC press, 2010. ISBN 9781439815915.

[11] K Jayasurya, G Fung, S Yu, Cary Dehing-Oberije, Dirk De Ruysscher, A Hope, Wilfried De Neve, Yolande Lievens, P Lambin, and ALAJ Dekker. Comparison of bayesian

network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. *Medical physics*, 37:1401–1407, 2010. doi: 10.1118/1.3352709.

[12] M Berkan Sesen, Ann E Nicholson, Rene Banares-Alcantara, Timor Kadir, and Michael Brady. Bayesian networks for clinical decision support in lung cancer care. *PloS one*, 8, 2013. doi: 10.1371/journal.pone.0082349.

[13] Arthur Jochems, Timo M Deist, Issam El Naqa, Marc Kessler, Chuck Mayo, Jackson Reeves, Shruti Jolly, Martha Matuszak, Randall Ten Haken, Johan van Soest, et al. Developing and validating a survival prediction model for nsclc patients through distributed learning across 3 countries. *International Journal of Radiation Oncology* Biology* Physics*, 99:344–352, 2017. doi: 10.1016/j.ijrobp.2017.04.021.

[14] Judea Pearl. Introduction to probabilities, graphs, and causal models. *Causality: models, reasoning and inference*, pages 1–40, 2000. doi: https://doi.org/10.1017/CBO9780511803161.003.

[15] Luis Torgo. *Data mining with R: learning with case studies*. Chapman and Hall/CRC, 2011. ISBN 9780367573980.

[16] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

[17] Marco Scutari. Learning bayesian networks with the bnlearn R package. *Journal of Statistical Software*, 35:1–22, 2010. doi: 10.18637/jss.v035.i03.

[18] Marek J Druzdzel. Smile: Structural modeling, inference, and learning engine and genie: a development environment for graphical decision-theoretic models. In *Aaai/Iaai*, pages 902–903, 1999. ISBN 0-262-51106-1.

[19] Richard E Neapolitan et al. *Learning bayesian networks*, volume 38. Pearson Prentice Hall Upper Saddle River, NJ, 2004. ISBN 978-0130125347.

[20] José A Gámez, Juan L Mateo, and José M Puerta. A fast hill-climbing algorithm for bayesian networks structure learning. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 585–597. Springer, 2007. doi: 10.1007/978-3-540-75256-1_52.

[21] Peter Lucas. *Bayesian networks in medicine: a model-based approach to medical decision making*. na, 2001.

[22] Vincenzo Valentini, Ruud GPM Van Stiphout, Guido Lammering, Maria Antonietta Gambacorta, Maria Cristina Barba, Marek Bebenek, Franck Bonnetain, Jean-Francois Bosset, Krzysztof Bujko, Luca Cionini, et al. Nomograms for predicting local recurrence, distant metastases, and overall survival for patients with locally advanced rectal cancer on the basis of european randomized clinical trials. *Journal of clinical oncology*, 29: 3163–3172, 2011. doi: 10.1200/JCO.2010.33.1595.

[23] W Van Gijn, Rgpm Van Stiphout, CJH Van De Velde, Vincenzo Valentini, G Lammering, Maria Antonietta Gambacorta, Lars Påhlman, K Bujko, and P Lambin. Nomograms to predict survival and the risk for developing local or distant recurrence in patients with rectal cancer treated with optional short-term radiotherapy. *Annals of Oncology*, 26:928–935, 2015. doi: 10.1093/annonc/mdv023.

[24] Koya Hida, Ryosuke Okamura, Soo Yeun Park, Tatsuto Nishigori, Ryo Takahashi, Kenji Kawada, Gyu-Seog Choi, and Yoshiharu Sakai. A new prediction model for local recurrence after curative rectal cancer surgery: development and validation as an asian collaborative study. *Diseases of the Colon & Rectum*, 60:1168–1174, 2017. doi: 10.1097/DCR.0000000000000896.

[25] Waad Farhat, Mohamed Azzaza, Abdelkader Mizouni, Houssem Ammar, Mahdi ben Ltaifa, Sami Lagha, Mohamed Kahloul, Rahul Gupta, Mohamed Ben Mabrouk, and Ali Ben Ali. Factors predicting recurrence after curative resection for rectal cancer: a 16-year study. *World journal of surgical oncology*, 17:173, 2019. doi: 10.1186/s12957-019-1718-1.

[26] Junjie Peng, Ying Ding, Shanshan Tu, Debing Shi, Liang Sun, Xinxiang Li, Hongbin Wu, and Sanjun Cai. Prognostic nomograms for predicting survival and distant metastases in locally advanced rectal cancers. *PLoS One*, 9:e106344, 2014. doi: 10.1371/journal.pone.0106344.

[27] Satish K Warrier, Joseph Cherng Kong, Glen R Guerra, Timothy J Chittleborough, Arun Naik, Robert G Ramsay, A Craig Lynch, and Alexander G Heriot. Risk factors associated with circumferential resection margin positivity in rectal cancer: a binational registry study. *Diseases of the Colon & Rectum*, 61:433–440, 2018. doi: 10.1097/DCR.0000000000001026.

[28] Qi Liu, Dakui Luo, Sanjun Cai, Qingguo Li, and Xinxiang Li. Circumferential resection margin as a prognostic factor after rectal cancer surgery: A large population-based retrospective study. *Cancer medicine*, 7:3673–3681, 2018. doi: https://doi.org/10.1002/cam4.1662.

CHAPTER 9

BAYESIAN NETWORK STRUCTURE FOR PREDICTING TWO-YEAR SURVIVAL

Authors:
Biche Osong, *Andre Dekker, Leonard Wee, Johan van Soest, Inigo Bermejo*

Submitted to: *SN Computer Science*

Abstract

Purpose
The aim of this study was to develop and internally validate a clinically plausible Bayesian network structure to predict two-year survival in patients diagnosed with non-small cell lung cancer (NSCLC) and primarily treated with (chemo) radiation therapy by combining expert knowledge and a learning algorithm.

Summary of background
The incidence of lung cancer has been increasing. Healthcare providers are trying to acquire more knowledge of the disease's biology to treat their patients better. However, the information available is more than humans can efficiently process. Predictive models such as Bayesian networks, which can intricately represent causal relations between variables, are suitable structures to model this information. However, commonly known methods for developing Bayesian network structures are limited in healthcare.

Patients and Methods
545 NSCLC patients treated primarily with (chemo) radiation therapy from Maastro clinic in the Netherlands between 2010 to 2013 were considered to develop this Bayesian network structure. All continuous variables were discretized before analysis. Patients with missing survival status and variables with more than 25% missing information were excluded. The causal relationships (arcs) between variables in the data were determined using the hill-climbing algorithm with domain experts' restrictions. The learning algorithm was run on a number of bootstrapped samples (B=400) and for the final structure, we kept the arcs present in at least 70% of the learned structures. Performance was assessed by computing the area under the curve (AUC) values and producing calibration plots based on a 5-fold cross-validation. In addition, an adapted pre-specified expert structure was compared with a structure developed from the method in this study.

Results
Tumor load was included in the main structure due to its high percentage (37%) of missingness and lack of added value. The final cohort used to develop the structure was reduced to 499, excluding 46 (8.4%) patients with missing survival status. The resulting structure's mean AUC and confidence interval to predict two-year survival was 0.614 (0.499 - 0.730). The AUC of the compared structures was only slightly above the chance level, but the structure based on the method in this study was clinically more plausible.

Conclusion
The results of this study shows that Bayesian network structures which combine expert knowledge with a rigorous structure learning algorithm produce a clinically plausible structure with optimal performance.

Introduction

Lung cancer is the second most common cancer and the leading cause of cancer morbidity and mortality in men and second for women after breast cancer worldwide, with non-small cell lung cancer (NSCLC) accounting for approximately 85% of this disease (1, 2). These statistics actuate healthcare providers to acquire more knowledge and understand the patient condition and disease characteristics for better patient management and treatment outcomes.

However, the amount of information that needs to be processed to ascertain if a patient will experience an event of interest can be challenging even for domain experts (3, 4). Furthermore, it has been shown that even experienced domain experts specialized in the treatment of lung cancer have limited to no capabilities for predicting patients' outcomes vis-a-vis prediction models (5–8). Predictive models such as Bayesian networks (BN), which can structurally represent a domain of interest by causally mapping the domain's variables, may be more suitable for modeling such information.

Bayesian network structures are either expert(s) specified or algorithm-based (9–11). However, these methods are limited in a clinical setting due to implausible casual relationships for algorithmic structures or bias for expert structures based on their experience and domain knowledge. Our prior work, which compared the performance of structures from these two sources, showed that algorithm-based structures perform relatively better but with little or no clinical interpretability (12). On the other hand, expert structures are more clinically interpretable but with relatively inferior performance. Therefore, this study aims to develop a Bayesian network structure that stems from both methods to predict two-year survival in patients diagnosed with lung cancer primarily treated with radiation therapy.

We hypothesize that a symbiotic relationship between domain experts and a robust learning algorithm (expert-algorithm) would yield a clinically interpretable and plausible Bayesian network structure to predict two-year survival for lung cancer patients with optimal performance.

Materials and methods

We retrospectively collected data of 545 non-small cell lung cancer (NSCLC) patients diagnosed between 2010 to 2013 and eligible for (chemo) radiotherapy treatment at Maastro Clinic, Maastricht, The Netherlands. Patients' demographics and clinical information such as age, gender, WHO performance status, TNM stage, tumor load, FEV, smoking status, chemotherapy type, and two-year survival status were extracted to establish the Bayesian network.

Bayesian network

Bayesian networks model the relationships between a set of variables. These relationships are represented in a directed acyclic graph (DAG), where each node in the graph signifies a variable (9).The direction of the link between nodes represents the influence dependency from the causal variable known as the parent node to the affected variable called the child node. Therefore, each variable can be a child or parent to numerous variables, but the process should not contain any loops. In other words, tracing the parent-to-child link should not connect a variable with itself or a variable functioning as a child and parent to another variable (11). The conditional probability table (CPT) represents the probabilities of each possible state of a node, given the states, its parent node may take (9, 10, 13).

Structure learning

The structure learning process was bootstrapped (B=400) with varying sample sizes at each run using the hill-climbing algorithm to identify the causal relationships (arcs) between variables in the dataset. Arc strength was evaluated as the rate of occurrence over all the bootstrap runs, and only arcs with an occurrence rate above 70% were included in the final structure. Domain knowledge from multiple experts in the field was employed to restrict the algorithm from forming arcs in clinically implausible directions (so-called blacklist) like age having a causal influence on gender (Table S9.1 in the supplementary materials). The pseudocode for the expert-algorithm method is outlined in algorithm 1.

Algorithm 1: Expert-Algorithm pseudocode

Input: Data **D**, Bootstrap run **B**, Restriction **R**, threshold ϵ

Output: Arcs as list **A**, BN

Step 1: Bootstrap Learning `// Learning on the bootstrap samples` d_i

for $i : 1...\boldsymbol{B}$ *in* $\boldsymbol{d}_i$ **do**

  A_j = hc($\mathbf{d}_i$, **B**, **R**) `// R: Expert restricted arcs`

end

Step 2: Arc Strength

$\mathbf{A} = \frac{\sum_{(j=1)}^{n} A_j}{\mathbf{B}}$ `// Count each arc in all bootstrap runs`

Step 3: Arc Thresholding

foreach arc in **A**

if *The arc strength* $\geq \epsilon$ `// Arc selection for final BN structure`

then

  Include in the final structure.

end

end

return BN

hc = Hill-climbing learning algorithm, **BN** = Bayesian Network

Statistical analysis

All analysis was conducted in R version 4.1.0 (14) using the bnlearn package (15) and GeNIe a Graphical Network Interface application (16) was used to visualize the developed Bayesian network structures. Tumorload and age were categorized into three groups with cutoff values at the 25th and 75th percentile and the force expiratory volume (FEV) was categorized based on experts opinions. Missing values were imputed using the Multivariate Imputation via Chained Equations (MICE) package (17). Patients with missing survival status and variables with more than 25% missingness were excluded from the analysis. The predictive performance of the resulting Bayesian network structure was assessed by computing the area under the curve (AUC) using a 5-fold cross validation technique and generating calibration plots.

The main structure was updated with the excluded variable using the **structural.em** function in the bnlearn package to check if the excluded variables having above 25% missingness were crucial. The function learns a Bayesian network from a dataset containing missing information by first inputting the missing data using the expectation-maximization (EM) algorithm and then finds the best possible structure based on the imputed data. The arcs of the main structure were used as a whitelist in the structure update process.

Results

Of the 545 patients in this study, 46 (8.4%) with a missing two-year survival status were excluded from this analysis, reducing the cohort to 499. Tumorload was excluded from further analysis due to its high percentage (37%) of missing information. The median age of patients in this study was 68 (33 - 89). Most of the patients in this study are ex-smokers, with the number of males almost twice that of females. Table 9.1 shows detailed descriptions of patients' characteristics for this study cohort.

The variable age was discretized into three groups with cutoff values at the 25th and 75th percentile (Figure S9.2 supplementary materials). Patients between the cutoff values are considered elderly as shown in equation 9.1 while patients below and above the cutoff values were considered adults and seniors, respectively. The forced expiratory volume (FEV) was also discretized into four groups based on experts' suggestions as shown in equation 9.2.

$$Age = \begin{cases} \text{(Below 61 years)} & - & \text{Adults} \\ \text{(61 - 76 years)} & - & \text{Elderly} \\ \text{(Above 76 years)} & - & \text{Seniors} \end{cases} \quad (9.1)$$

$$FEV = \begin{cases} \text{(Below 40)} & - & \text{Severe} \\ \text{(40 - 59)} & - & \text{Moderate} \\ \text{(60 - 79)} & - & \text{Mild} \\ \text{(Above 79)} & - & \text{Normal} \end{cases} \quad (9.2)$$

The WHO performance status was recategorized into four groups by combining patients in the limited (3) and bed-bound (4) categories into the same category (bed-bound) because of the very low number of patients in the two categories (Table 9.1). Also, they both have similar characteristics (See Table S9.2 in the supplemental material for further explanations).

Diagnostic plots were created to ensure that the imputations have converged to the desired distribution. The convergence check plot (Figure S9.1 supplemental material) of the imputed values suggests the imputation has converged to the target distribution. Furthermore, the density plot (Figure S9.4 supplemental material) which compares the distribution of the imputed and observed values confirms that the imputations are reasonable since the distribution of the imputed and observed values are very similar.

Table 9.1. Overview of patient demographics and clinical characteristics

Variable	Levels	Two years survival		Total
		False	True	
Age at RT	Mean (sd)	68 (10.9)	67 (10.1)	68 (10.5)
FEV	Mean (sd)	76 (20.8)	78 (22.4)	77 (21.5)
	Missing	68 (22.5%)	42 (21.3%)	110 (22.0%)
Tumorload	Mean (sd)	136 (143.6)	92 (150.7)	113 (145.7)
	Missing	111 (36.8%)	64 (32.5%)	204 (37.4%)
Sex	Male	205 (63.1%)	120 (36.9%)	325 (65.1%)
	Female	97 (55.7%)	77 (36.9%)	174 (34.9%)
Metastasis	No	252 (60.3%)	166 (39.7%)	418 (90.9%)
	Yes	28 (66.7%)	14 (33.3%)	42 (09.1%)
	Missing	22 (56.4%)	17 (43.6%)	39 (07.8%)
Chemotherapy type	No chemo	71 (66.4%)	36 (33.6%)	107 (25.6%)
	Sequential	31 (60.8%)	20 (39.2%)	51 (12.2%)
	Concurrent	151 (58.1%)	109 (41.9%)	260 (62.2%)
	Missing	49 (60.5%)	32 (39.5%)	81 (16.2%)
Smoking status	Non smokers	07 (50.0%)	07 (50.0%)	14 (03.1%)
	Quit smoking	194 (62.4%)	117 (37.6%)	311 (68.2%)
	Smokers	76 (58.0%)	55 (42.0%)	131 (28.7%)
	Missing	25 (58.1%)	18 (41.9%)	43 (08.6%)
WHO performance status	Active (0)	46 (52.9%)	41 (47.1%)	87 (17.7%)
	Restricted (1)	172 (61.4%)	108 (38.6%)	280 (57.0%)
	Self care (2)	65 (64.4%)	36 (35.6%)	101 (20.6%)
	limited (3)	13 (65.0%)	07 (35.0%)	20 (04.1%)
	Bedbound (4)	02 (66.7%)	01 (33.3%)	03 (00.6%)
	Missing	04 (50.0%)	04 (50.0%)	08 (01.6%)
Tumor stage	T0	45 (61.6%)	28 (38.4%)	73 (15.7%)
	T1	89 (64.0%)	50 (36.0%)	139 (29.8%)
	T2	50 (60.2%)	33 (39.8%)	83 (17.8%)
	T3	100 (58.5%)	71 (41.5%)	171 (36.7%)
	Missing	18 (54.5%)	15 (45.5%)	33 (06.6%)
Nodal stage	N0	77 (58.3%)	55 (41.7%)	132 (27.6%)
	N1	16 (51.6%)	15 (48.4%)	31 (06.5%)
	N2	114 (58.5%)	81 (41.5%)	195 (40.8%)
	N3	85 (70.8%)	35 (29.2%)	120 (25.1%)
	Missing	10 (47.6%)	11 (52.4%)	21 (04.2%)
Two year survival	False	-	-	302 (60.5%)
	True	-	-	197 (39.5%)
	Missing	-	-	46 (08.4%)

RT: Radiotherapy, **Chemo**: Chemotherapy, **WHO**: World Health Organization, **FEV**: Forced expiratory volume

Table 9.2 shows the results of the aggregated arcs from the bootstrap structure learning process using the hill-climbing (hc) algorithm and expert restriction (Blacklist). The arc strength shows the percentage of occurrence of each arc in the bootstrap runs.

Table 9.2. Bootstrap run output with arcs and strength for the main structure.

Arcs Dependencies (A)		Strength
Parent node	**Child node**	
Gender	Two year survival	0.65
Gender	FEV	0.64
WHO PS	**Two year survival**	**0.70**
WHO PS	Chemotherapy type	0.43
Age	Two year survival	0.61
Age	**WHO PS**	**0.84**
Age	**Chemotherapy type**	**0.99**
Tumor stage	**Two year survival**	**0.73**
Tumor stage	WHO PS	0.43
Tumor stage	**Nodal stage**	**0.91**
Tumor stage	FEV	0.64
Tumor stage	**Chemotherapy type**	**0.94**
Nodal stage	**Two year survival**	**0.84**
Nodal stage	**WHO PS**	**0.91**
Nodal stage	**FEV**	**1.00**
Nodal stage	**Chemotherapy type**	**1.00**
Metastasis	Two year survival	0.42
Metastasis	**WHO PS**	**0.92**
Metastasis	FEV	0.66
Metastasis	**Chemotherapy type**	**0.76**
FEV	Two year survival	0.63
FEV	**WHO PS**	**0.86**
FEV	**Chemotherapy type**	**0.97**
Chemotherapy type	Two year survival	0.34
Smoking status	**Two year survival**	**0.86**
Smoking status	**FEV**	**0.82**

FEV: Forced expiratory volume, **WHO PS**: World Health Organization performance status

A threshold of 0.7 was chosen to decide which arcs should be included or excluded in the Bayesian network structure. A higher threshold value ensures that the conditional probability table (CPT) of the outcome does not grow too large, which can cause the structure to overfit. Therefore, the chosen threshold helps restrain the structure from overfitting but allows enough room for structural complexity for optimal performance.

The resulting Bayesian network structure based on the chosen threshold is presented in figure 9.1. Of the 26 arcs produced during the bootstrap structure learning process, only 15 had an arc strength above the selected threshold and were used to develop the Bayesian network structure. Four variables directly influenced the response of interest (gray arcs), while gender was isolated in the structure because it was neither a parent nor a child to any variable in the network.

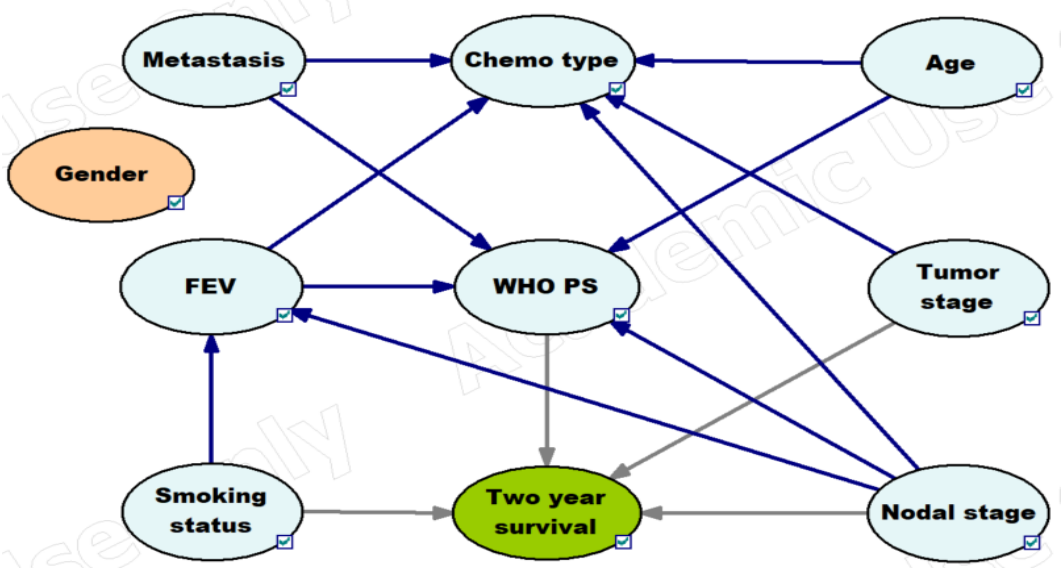

Figure 9.1. *Resulting Bayesian network structure to predict two-year survival from the expert-algorithm method. The oval structure represents the variables (Node), and the arrows indicate the direction of the causal-effect relationships. Grey arrows indicate a direct parental link to the outcome of interest*

***FEV** = Forced expiratory volume, **Chemo** = Chemotherapy, **WHO PS** = World Health Organization performance status*

Figure 9.2 shows the performance assessment results when the resulting Bayesian network structure was used to predict two years survival in lung cancer patients using the repeated (r=50) 5-fold cross-validation technique. The left figure shows the area under the curve and confidence intervals of the respective folds with a mean value of 0.614 (0.499 - 0.730). The right figure gives a measure of how similar the predicted probabilities are to the observed probabilities with calibration assessed in terms of the degree of deviation of the points (color) from the 45-degree line (dotted gray).

Updating the main structure with the excluded variable included three arcs (red) between **tumor load** and **tumor stage**, **tumor load** and the presence of **metastasis**, the presence of **metastasis** and **two years survival** (Figure S9.3 in the supplemental material). However, the addition of these arcs had no significant improvement on the mean AUC value. Gender was again not connected to any other variable in the structure.

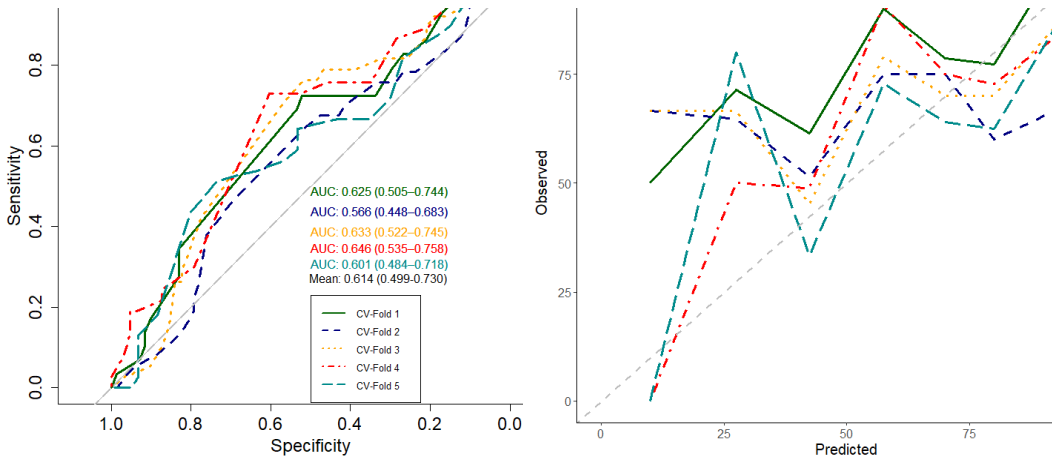

Figure 9.2. *The area under the curve and calibration plot of the structure for predicting two-year survival.*

To assess the expert-algorithm method, the structure and performance of an expert pre-specified structure from Jochems et al. (18) was compared with a structure resulting from the expert-algorithm procedure with a threshold of 0.7 and the same variables. The aggregated arcs results from the bootstrap runs based on these variables are presented in table 9.3

Table 9.3. Bootstrap run output with arcs and strength for structure comparison.

Arcs Dependencies (A)		Strength
Parent node	**Child node**	
WHO PS	Two year survival	0.43
Age	Two year survival	0.38
Age	**WHO PS**	0.94
Tumor stage	Two year survival	0.36
Tumor stage	WHO PS	0.10
Tumor stage	Nodal stage	0.68
Nodal stage	**Two year survival**	0.74
Nodal stage	**WHO PS**	0.92

WHO PS: World Health Organization performance status

The structure from Jochems et al. (18) was adapted in this analysis because of the missing total tumor dose variable. Figure 9.3 shows the adapted structure from Jochems et al. (18) and that resulting from the expert-algorithm method respectively. The only similarity in the structures is that the outcome has just one parent but the variables are completely different with the adapted structure having WHO performance score as parent and nodal stage for the expert-algorithm structure. Based on the differences, the expert-algorithm structure uses one variable and arc less with the outcome being an end node (having no child) compared to the adapted structure.

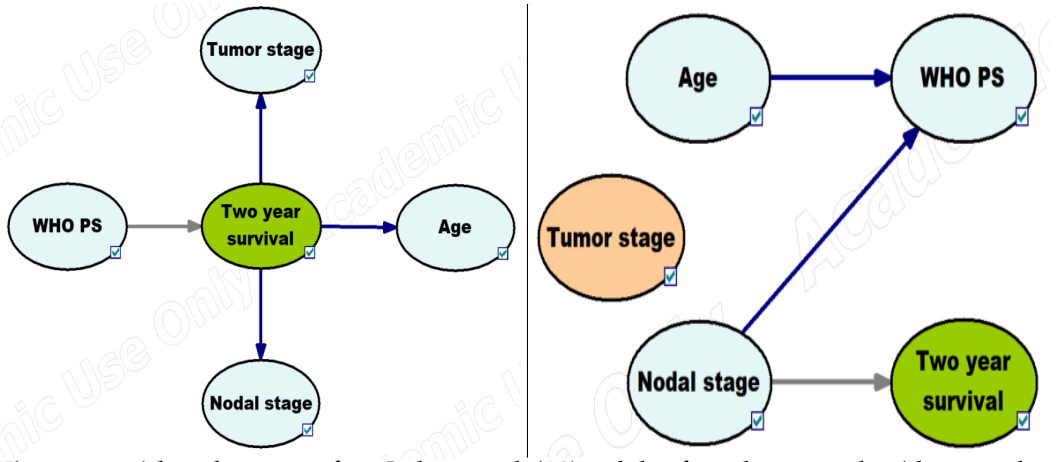

Figure 9.3. *Adapted structure from Jochems et al. (18) and that from the expert-algorithm procedure respectively. The oval structure represent the variables (Node), and the arrows indicate the direction of the causal-effect relationships. Grey arrows indicate a direct parental link to the outcome of interest*
***WHO PS**: World Health Organization performance status*

The performance of both structures was only slightly better than flipping a coin (Figure S9.5 supplemental material) with an area under the curve of 0.56 (0.517 - 0.613) for the expert-algorithm structure and 0.53 (0.489 - 0.582) for the adapted structure from Jochems et al. (18). Though both structures had poor performance with just one arc to the outcome, the expert-algorithm structure had a slightly higher discriminating ability than the adapted structure, but this difference was not statistically significant (p-value = 0.413).

To further evaluate the performance of the structures, their respective calibration plots were produced and overlaid (Figure S9.5 supplemental material). They show that both plots were poorly calibrated given how distant the points are from the diagonal dotted gray line. However, the expert-algorithm structure better calibrated relatively with more points closer to the diagonal line.

Discussion

We have developed a novel structure learning method to produce clinically plausible and interpretable Bayesian networks resulting from an interplay of a learning algorithm and experts' restrictions. The structure produced due to our expert-algorithm method was used to predict two-year survival in NSCLC patients. A resulting structure from the application of our expert-algorithm method was compared to an adapted expert pre-specified structure. Both structures produced comparable results in terms of AUC and calibration. However, the expert-algorithm structure had a slightly better performance with one variable and arc less than the adapted structure and more clinically plausible arcs.

Numerous Bayesian network structures have been developed to predict survival in lung cancer patients (19), and some of these structures stem from our group (12, 18, 20–22). These structures were either inferred from the data by a structure learning algorithm or pre-specified by expert(s). Jochems et al. (18) has even compared the performance of structures derived from both methods and showed that expert-based structures were more performant than algorithm-based structures, although the difference was not statistically significant. To our knowledge, this is the first time a Bayesian network structure has been developed from routine clinical care data that applies both structure learning methods while considering the clinical sanity and interpretability of the resulting structure—in other words, developing a Bayesian network structure which is suited for clinical implementation, because Bayesian network structures which captures domain knowledge are more interpretable which is essential for clinical decision making.

Bayesian network structure learning process can be time-consuming and computer-intensive, especially for expert base and algorithmic structures like the hill-climbing algorithm, which test pairs of variables to determine whether edges should be included or removed from the structure respectively. Therefore applying our expert-algorithm method, which restricts candidate solutions (arcs) from being evaluated during the search process, could significantly improve structure learning in radiotherapy fields, which involves using high dimensional data, as is the case with radiomics studies. Furthermore, our expert-algorithm method could serve as a means to perform variable selection in the structure learning process of Bayesian networks, something missing in literature since variable selection is mainly performed manually by experts. When a high threshold is applied to the bootstrap object, it leads to the removal of arcs with strength smaller than the specified threshold and possibly variables with arc strength inferior to the threshold as in the case for gender in figure 9.1. Finally, the expert restriction and the use of threshold ensure that the resulting structure is clinically correct and includes only relevant relationships in the data with optimal performance.

This study hypothesized that combining experts' knowledge with a learning algorithm would yield a more clinically correct structure. The results comparing the structures from both methods support this hypothesis. Although both structures performed relatively poorly, we observed a significant improvement in our expert-algorithm structure as most of the arcs in the adapted structure were either reversed or modified to clinically acceptable arcs. However, the arc from WHO PS to the outcome, present in the adapted structure, was absent in the expert-algorithm structure. One possible explanation for this difference could be that WHO PS alone is not a good correlate to survival, but its interplay with other variables increases its influence on survival, as seen in the primary structure (Figure 9.1) with an arc strength of 0.73 as opposed to 0.43 (Table 9.2 and 9.3 respectively) . Similar conclusions can be drawn from the study by Jayasurya et al. (20) as the arc (WHO PS to Survival) was also present in their structure with a link weight of -0.169, suggesting that WHO PS has limited ability to predict survival correctly.

The structures presented in this study are not intended for clinical use at the moment but to encourage further research on how to merge these two methods best to develop a Bayesian network structure for a more clinically correct structure but with optimal predictive performance. Therefore, this study is not devoid of limitations mainly because of its retrospective nature, which implies the possibility of data bias. Furthermore, the main structure did not include predictive variables such as the number of positive lymph nodes on the PET scan (PLNS) and tumor load. Although tumor load was available as a variable in this study, it was excluded due to missing information. Faehling et al. (23) has shown that tumor load is an essential factor for overall survival, and patients with lower tumor load have a better outcome than patients with larger tumor load. PLNS, on the other hand, was unavailable in this study which makes the structural comparison somewhat unfair. Also, information is lost with the discretization of continuous variables, which might explain the low predictive power of tumor load in this study coupled with the high missingness. Lastly, this study's threshold selection is set arbitrarily, which means a large threshold will lead to a sparse network that only partially represents the domain, and a small threshold yields a dense network (complex, squiggly, and harder to read) which might overfit the data. Future researchers should focus on finding an optimal threshold that addresses the shortcomings of the present thresholding.

Conclusion

We have developed a Bayesian network structure from routine clinical data for predicting two-year survival in lung cancer patients treated with (chemo) radiotherapy. Our expert-algorithm method uses bootstrapping with arc restriction in the structure learning process and assesses the robustness of causal relationships. Therefore, selecting the most robust relationships overall bootstrapping samples produces a structure that captures all relevant relationships within the data with a reduced chance of adding spurious links. In the future, we intend to use this method to evaluate different structures learned from different data sets and perform a privacy-preserving distributed learning approach to structure learning in Bayesian networks.

Supplemental materials

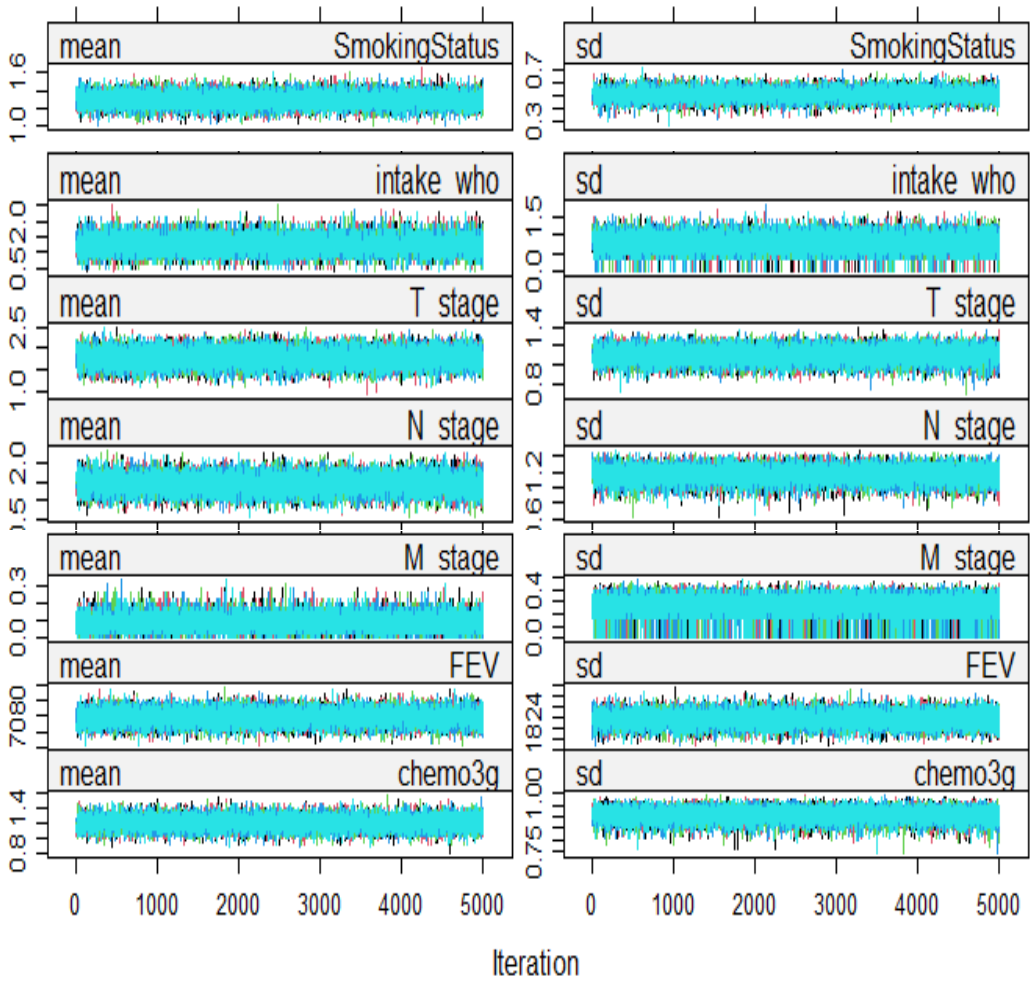

Figure 9.1. *The convergence check plot shows the mean (left) and standard deviation (right) of the imputed values against iteration number. The plot suggests the imputation has converged to the target distribution given the good mix/intermingling of the streams and its trends-free nature at the later part of the iterations*

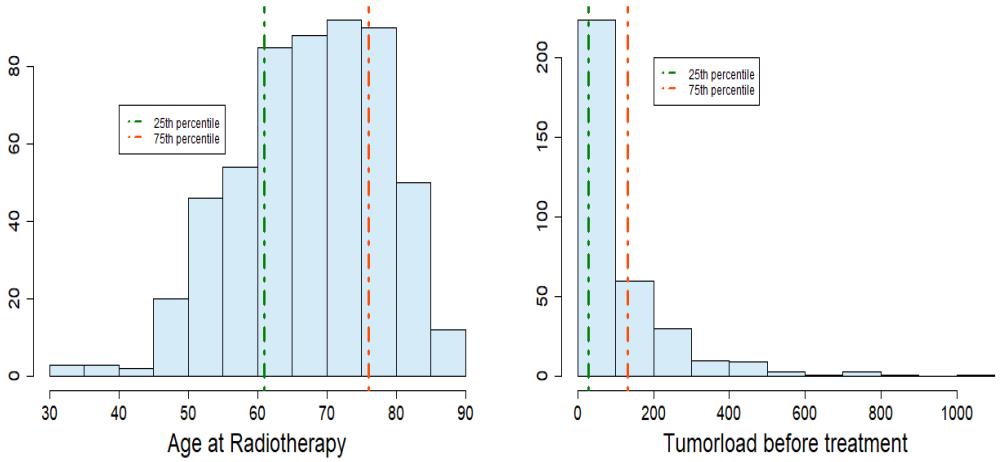

Figure 9.2. *Histogram of age and tumorload with vertical lines at the 25th and 75th percentile respectively.*

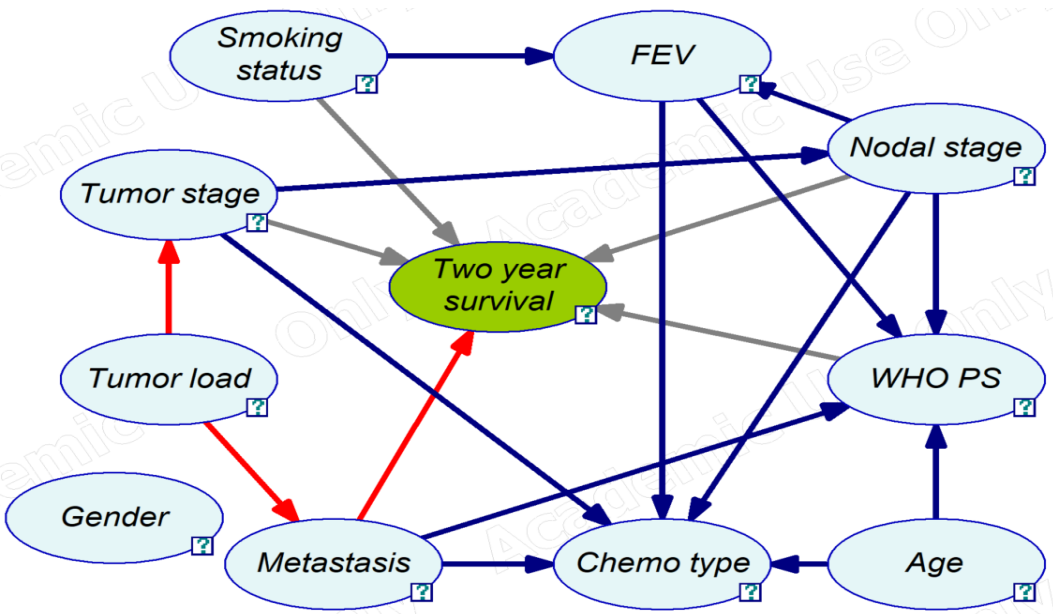

Figure 9.3. *Updated Bayesian network structure to predict two-year survival with the structural.em function in the bnlearn package. The oval structure represent the variables (Node), and the arrows indicate the direction of the causal-effect relationships. Grey arrows indicate a direct parental link to the outcome of interest and red arrows indicates additional arcs.*
***FEV**: Forced expiratory volume, **Chemo**: Chemotherapy, **WHO PS**: World Health Organization performance status*

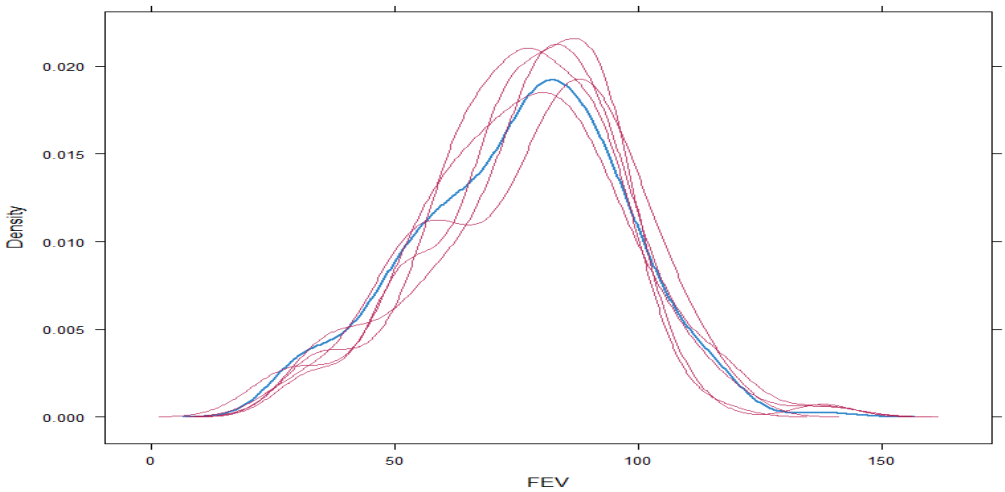

Figure 9.4. *Kernel density estimates for the observed data (blue) and the imputed data (thin red lines) for forced expiratory volume (FEV).*

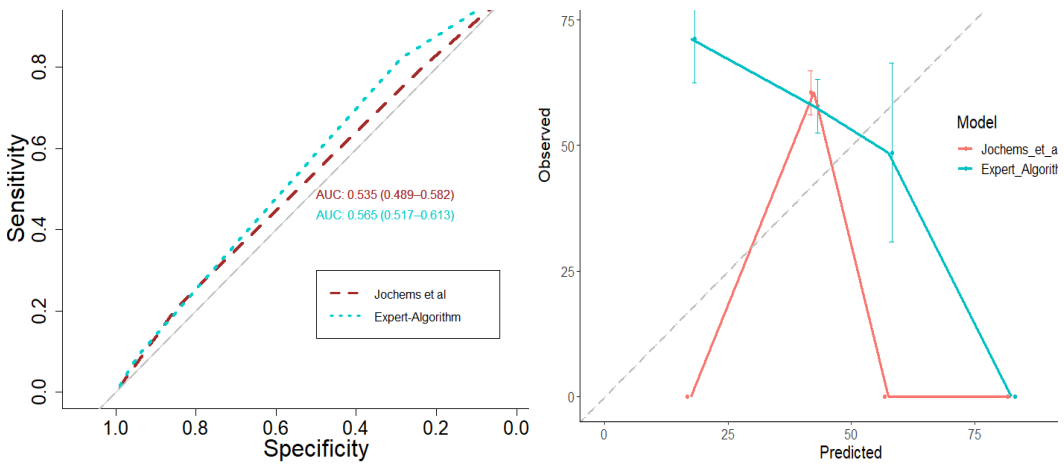

Figure 9.5. *Performance of the adapted and expert-algorithm structure in terms of the area under the curve and calibration respectively.*

Table 9.1. Blacklist restriction of patient characteristics.

	Gender	WHO PS	Age	Chemo type	Tumorload	Smoking status	Tumor stage	Nodal stage	Metastasis	FEV	Two year survival
Gender	O	X	X	X	X	X	X	X	X	L	L
WHO PS	X	O	T	T	T	X	T	T	T	T	L
Age	X	L	O	L	X	X	X	X	X	X	L
Chemo type	X	L	T	O	T	X	T	T	T	T	L
Tumorload	X	L	X	L	O	X	L	L	L	L	L
Smoking status	X	X	X	X	X	O	X	X	X	X	L
Tumor stage	X	L	X	L	T	X	O	L	X	L	L
Nodal stage	X	L	X	L	T	X	T	O	X	L	L
Metastasis	X	L	X	L	T	X	X	X	O	L	L
FEV	T	L	X	L	T	X	T	T	T	O	L
Two year survival	T	T	T	T	T	T	T	T	T	T	O

FEV: Forced expiratory volume, **Chemo**: Chemotherapy
WHO PS: World Health Organization performance status
T: Top determines left, **L:** Left determines top, **X:** Both direction blocked, **O:** Diagonal

Table 9.2. WHO Performance status

Grade	Explanation of activity
0	Fully active, able to carry on all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up and about more than 50% of waking hours
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair.
5	Dead

Bibliography

[1] Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.

[2] Wenpeng You and Maciej Henneberg. Cancer incidence increasing globally: the role of relaxed natural selection. *Evolutionary applications*, 11(2):140–152, 2018.

[3] Cary Oberije, Georgi Nalbantov, Andre Dekker, Liesbeth Boersma, Jacques Borger, Bart Reymen, Angela van Baardwijk, Rinus Wanders, Dirk De Ruysscher, Ewout Steyerberg, et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. *Radiotherapy and Oncology*, 112(1):37–43, 2014.

[4] Amy P Abernethy, Lynn M Etheredge, Patricia A Ganz, Paul Wallace, Robert R German, Chalapathy Neti, Peter B Bach, and Sharon B Murphy. Rapid-learning system for cancer care. *Journal of Clinical Oncology*, 28(27):4268, 2010.

[5] Cary Oberije, Georgi Nalbantov, Andre Dekker, Liesbeth Boersma, Jacques Borger, Bart Reymen, Angela van Baardwijk, Rinus Wanders, Dirk De Ruysscher, Ewout Steyerberg, et al. A prospective study comparing the predictions of doctors versus models for treatment outcome of lung cancer patients: a step toward individualized care and shared decision making. *Radiotherapy and Oncology*, 112(1):37–43, 2014.

[6] Crealy Louise. Algorithm performs better than doctors at predicting mental health outcomes, 2018. URL https://www.abc.net.au/news/2018-09-27/ai-better-than-doctors-at-predicting-mental-health-outcomes/10311388. Accessed: 2019-02-20.

[7] Montague Jules. Why doctors get it wrong about when you will die, 2015. URL https://www.theguardian.com/lifeandstyle/2015/jun/02/doctors-predict-patient-die-prognosis-wrong. Accessed: 2019-02-20.

[8] Taylor Paul. Can doctors really predict how long one has to live?, 2016. URL https://healthydebate.ca/personal-health-navigator/can-doctors-really-predict-long-one-live. Accessed: 2019-02-20.

[9] Judea Pearl. Bayesian networks. 2011.

[10] Jose M Pena. Finding consensus bayesian network structures. *Journal of Artificial Intelligence Research*, 42:661–687, 2011.

[11] Marco Scutari, Catharina Elisabeth Graafland, and José Manuel Gutiérrez. Who learns better bayesian network structures. 2018.

[12] Biche Osong, Carlotta Masciocchi, Andrea Damiani, Inigo Bermejo, Elisa Meldolesi, Giuditta Chiloiro, Maaike Berbee, Seok Ho Lee, Andre Dekker, Vincenzo Valentini, et al. Bayesian network structure for predicting local tumor recurrence in rectal cancer patients treated with neoadjuvant chemoradiation followed by surgery. *Physics and imaging in radiation oncology*, 22:1–7, 2022.

[13] Marco Scutari and Jean-Baptiste Denis. *Bayesian networks: with examples in R*. CRC press, 2021.

[14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

[15] Marco Scutari. Package 'bnlearn'. *Bayesian network structure learning, parameter learning and inference, R package version 4.4*, 1, 2019.

[16] Marek J Druzdzel. Genie: A development environment for graphical decision-analytic models. In *Proceedings of the AMIA Symposium*, page 1206. American Medical Informatics Association, 1999.

[17] Stef van Buuren, Karin Groothuis-Oudshoorn, Alexander Robitzsch, Gerko Vink, Lisa Doove, and Shahab Jolani. Package 'mice'. *Computer software]. Retrieved from: http://cran. r-project. org/web/packages/mice/mice. pdf*, 2015.

[18] Arthur Jochems, Timo M Deist, Issam El Naqa, Marc Kessler, Chuck Mayo, Jackson Reeves, Shruti Jolly, Martha Matuszak, Randall Ten Haken, Johan van Soest, et al. Developing and validating a survival prediction model for nsclc patients through distributed learning across 3 countries. *International Journal of Radiation Oncology* Biology* Physics*, 99(2):344–352, 2017.

[19] M Berkan Sesen, Ann E Nicholson, Rene Banares-Alcantara, Timor Kadir, and Michael Brady. Bayesian networks for clinical decision support in lung cancer care. *PloS one*, 8 (12):e82349, 2013.

[20] K Jayasurya, G Fung, S Yu, Cary Dehing-Oberije, Dirk De Ruysscher, A Hope, Wilfried De Neve, Yolande Lievens, P Lambin, and ALAJ Dekker. Comparison of bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. *Medical physics*, 37(4):1401–1407, 2010.

[21] Andre Dekker, Andrew Hope, Philippe Lambin, and Patricia Lindsay. Survival prediction with bayesian networks in more than 6000 non-small cell lung cancer patients. *medRxiv*, 2021.

[22] Melle S Sieswerda, Inigo Bermejo, Gijs Geleijnse, Mieke J Aarts, Valery EPP Lemmens, Dirk De Ruysscher, André LAJ Dekker, and Xander AAM Verbeek. Predicting lung cancer survival using probabilistic reclassification of tnm editions with a bayesian network. *JCO clinical cancer informatics*, 4:436–443, 2020.

[23] M Faehling, B Schwenk, M Kopp, S Fallscheer, S Kramberg, and R Eckert. Immuno-oncological treatment and tumor load in non-small cell lung cancer (nsclc): Case-control analysis of overall survival (os) in routine clinical practice. *Annals of Oncology*, 30:ii60, 2019.

Part VI
Discussion

CHAPTER 10

GENERAL DISCUSSION AND FUTURE PERSPECTIVES

Introduction

Machine learning methods that transform (Big) Data into knowledge have undergone unprecedented changes over the last decade, drawing much attention and interest in numerous life science fields. Their potential and impact on advancing healthcare and, more specifically, radiation oncology are vast. Unfortunately, many assume machine learning models are "black boxes" capable of magically providing the needed answers, with no interest in the reasoning behind the answer. The sad reality in medicine is that most of these developed prediction models will rarely be applied in real-world settings. In the model-building process, the two most pivotal things are the research question and the solution provided by the model. The end-users who implement the actionable insight generated by these trained prediction models might not know the intricate details of how the model works. However, since the model will be an integral part of the decision-making process, they have the right to pose the question, "how does the model arrive at its decisions?" Answers to such questions might help build trust between a machine-learning model and end-users.

To this end, this thesis tries to bridge the gap between clinical (Big) data and actionable insights using machine learning models to improve patient management and care. In particular, we focused on models representing how relevant features relate to each other and the outcomes of interest in the best, most straightforward manner possible. These relationships between variables and outcomes create a holistic view of the treatment process leveraged for informed clinical decisions. Furthermore, this thesis embedded the importance of including experts who are most often the end-users in the model-building process from a broader perspective to assist in making these models clinically meaningful.

The introduction section gives the skeletal content of this thesis and briefly talks about the disease, its treatment regarding radiation, the data it generates, and three fitted interpretable models. The theoretical section discusses the importance of Big Data in radiation oncology, from improving operational efficiency to clinical excellence for better cancer care and the challenges surrounding clinical Big Data. The analytical section, which is also the most extensive part of this thesis, uncovers how available patients' information (diagnostic, treatment, and observational information) is related to each other and a specific outcome of interest to predict the probability of future patients developing the outcome at a patient-specific level using three different machine learning methods. This chapter will briefly summarize the results of this thesis and discuss some challenges encountered during the model-building process and give some general recommendations. Then the importance of interpretable models and the value of experts for model development are discussed. Finally, some future perspectives are stated before this thesis is concluded.

Results summary

Big Data's potential in radiation oncology is tremendous, as discussed in the theory section in chapters 2 and 3. However, the high spread of data across many sources governed by different states, hospitals, and administrative bodies poses a significant challenge for Big Data applications in healthcare to reach their full potential, mainly due to patient privacy concerns and institutional reputation.

With the concept of data-driven medicine gradually being embraced as complementary to evidence-based medicine, the development of prediction models for various outcomes as implemented in the analysis section chapters 4 to 9 to assist caregivers with their decision-making tasks is needed. The regression part primarily focuses on the overall survival of patients for two different diseases. Chapter 4 looks at 3-years survival for women with cervical cancer and found that age, FIGO stage, and change in neutrophil-to-lymphocyte ratio ($\triangle$NLR) were significantly associated with survival. The next chapter found age, spinal cord compression, brain metastasis, visceral metastasis, WHO performance status, sex, and the primary tumor itself to be relevant for predicting 1, 3, and 6 months overall survival for patients whose primary tumor has metastasized to the spine. Model performance in this part based on the concordance index, and the 95% confidence interval was 0.73 (0.64 – 0.88) and 0.72 (0.68 – 0.75) for chapters 4 and 5, respectively. Their stratified Kaplan-Meier curves for the created risk groups were significantly separated with decent calibrations, suggesting the models have good discriminating power.

The decision tree developed for elderly cancer patients in chapter 6 showed that the hospitalization status of the patient is an essential determinant of radiotherapy compliance. The second most relevant predictor for outpatients was gender, while the comorbidity index was more important for outpatients. Other variables like performance status, cancer type, age, insurance status, fractionation type, and treatment aim were also part of the tree. These variables achieved a mean area under the curve (AUC) and a 95% confidence interval value of 0.71 (0.66 - 0.77). In chapter 7, the tree methodology was extended to time-to-event data to predict survival in cervical cancer patients. The tree had Figo staging as the primary determinant of survival, with age and tumor size also splitting to form other branches on the tree, with the SCC-Ag level being the last internal node. The tree had a concordance index of 0.78 (95% CI, 0.71 - 0.86) and 0.71 (95% CI, 0.63 - 0.89) in the training and external validation data, respectively. The Kaplan-Meier curve for the created risk group had a significant split in the training and validation cohort.

The last part of the analyses section is focused on the development and comparison of Bayesian network structures. The expert elicited Bayesian network structure developed in chapter 8 to predict tumor recurrence in rectal cancer patients was more clinically valid than the algorithm-based structure, which relies on retrospective data that might be biased to

build the Bayesian network structure. However, the structural performance of the expert-elicited structure was inferior to that of the algorithm-based structure. In chapter 9, a hybrid approach that uses experts' restriction of arcs (blacklist) and an algorithm's knowledge of the data (expert algorithm) was taken to develop the Bayesian network structures to predict two-year survival in patients diagnosed with non-small cell lung cancer. The mean area under the curve and confidence interval of the structure was 0.61 (0.49 - 0.73). The resulting structure based on the expert algorithm method outperformed an adapted expert pre-specified structure in terms of clinical correctness and area under the curve.

Model building challenges

Predictive models are fast becoming an integral component of healthcare since they can learn from the available patient information and provide insights that can assist patients and caregivers in making informed decisions. However, predictive modeling has several demanding and challenging tasks that could severely affect the model if not adequately handled in the development process.

Data quality

The first evident level of demand is the data quality, given that prediction models use data to learn how the variables are related to one another and the outcome of interest. Therefore the quality of data on which a model is developed determines its quality and trustworthiness. Given that trust is a critical aspect in healthcare, accuracy, completeness, consistency, credibility, and timeliness are used to judge data quality. Therefore, as the need to learn from each other's data via Big Data analytics to improve knowledge discovery and innovation continues to grow, ensuring data is of high quality in our respective silos is pivotal.

Missing information

Missing information is an inevitable problem in real-world datasets especially in routine clinical care as caregivers will not always have access to or time to note all desired data elements. (1, 2). However, when a dataset contains missing information, there is always a plausible reason for missingness (3), whether known or not. Therefore, understanding the possible cause(s) of the missing information and an appropriate method to handle the missingness is essential, especially for small sample size data.

Generally, most researchers handle missing information within their dataset using the listwise deletion method (4, 5), where all individuals within the data with missing information on any variable in the data are removed from further analyses (complete-case analysis). This method and its variant pairwise deletion (available-case analysis) are problematic in machine learning, since dropping observations could dramatically impact

the sample size. Reducing sample size leads to loss of information, reduced statistical power (6), less robust model (4), and potential bias in the estimated parameter (5). Single imputation methods such as zero imputation (7), mean imputation (8), hot deck imputation (9) and its algorithmic variant K-Nearest Neighbor imputation used to handle missing information at the data level have been largely criticized by researchers for their inability to properly account/handle the uncertainty induced by the imputed values (3).

In chapter 8, the missingness within the data was handled differently from the methods mentioned above. Here, the missing information is labeled to represent its state in the Bayesian network. Lin and Haug (10) have proven this method is superior to the complete-case analysis and single value imputation in terms of predictive power. The multiple imputation (MI) procedure that replaces each missing value with a set of imputed values is, however, the most commonly known herculean alternative method to handle missing information at the data level since it addresses most of the shortcomings of the single imputation methods (5, 9, 11). One way to roughly assess or ascertain the validity of the results obtained from the MI procedure is by a visual inspection of the trace (Chapter 9) and autocorrelation function plots (12, 13). The multiple imputation technique is implementable in most popular statistical tools like R, python, SAS, and SPSS.

The maximum likelihood (ML) and its superior iterative variant expectation-maximization (EM) (14) are better and more sophisticated techniques for handling missing information at the algorithmic level. The maximum likelihood method, which estimates the model's parameters directly by maximizing the incomplete data's likelihood function, is better than the multiple imputation method (15). The EM technique is much more popular with Bayesian network models (10, 16, 17), and it is implementable in some R packages (18, 19), while the other methods like MI are much more associated with regression models. In healthcare, we recommend using experts' predefined Bayesian network structure or learning one with experts' restriction(s) as in chapter 9 when applying the EM technique. These options ensure the structure stays in sync with the clinical process.

Decision trees, on the other hand, have an inbuilt mechanism that helps them deal with missing information in a more straightforward and less time-consuming manner by simply ignoring or labeling the missing information as a category during the splitting process (20–22).

Class imbalance

Class imbalance in binary responses, also called rare events, is a common datasets problem that critically affects the prediction of machine learning models in most domains. An imbalanced class dataset occurs when the number of events in one particular class of the binary response is much smaller than the other nonevents class. Building a prediction model where the response class distribution is imbalanced is very difficult. During the

learning process, the model implicitly assumes that the response classes are balanced (number of events and nonevents) and tries to optimize the overall classification accuracy based on this assumption. Therefore, during classification, the model favors the majority class, resulting in poor accuracy in detecting the minority class observations, which is often the class of interest (23, 24).

Researchers often try to deal with this imbalance class problem naturally by collecting a significant amount of observations to include as much data as possible of the rare events. However, this method can significantly increase the data collection demands and cost without guaranteeing enough events to alleviate the minority or rare class detection problem. The data in chapters 6 and 8 had a severe class imbalance problem even with a sample size of approximately five hundred and seven thousand observations respectively. The sampling technique, which deals with class imbalance problems at the data level by either oversampling the minority class or undersampling the majority class to achieve a balanced class distribution, is another option (25). This technique, however, has some drawbacks, first in defining an appropriate sampling ratio (26). Secondly, undersampling leads to loss of information, and the sampled group might not be an accurate representation of the actual population (27) while oversampling can lead to model overfitting (28). The two-step approach that first clusters the majority class and then uses the medoid (center of cluster) of each cluster with the rare events to train the model is a better alternative as it reduces the severity of population misrepresentation compared to the undersampling method, which randomly discards observations.

Methods such as cost-sensitive learning can deal with the imbalanced class distribution problem at the algorithm level by assigning a penalty call "cost" for every incorrect prediction or misclassification made by the algorithm. Therefore, setting a high cost to the misclassification of the minority class sample helps to improve the model performance without modifying the class distribution in the data (29, 30). Decision trees that aim to find logical rules that separate the observations within a dataset into different classes of homogeneous groups are among the algorithms most used in combination with the cost-sensitive learning technique for optimal performance (31–34).

Feature selection

In predictive modeling, as stated before, the aim is to find the relationship, if any exist, between the predictor variable(s) and the outcome of interest, and then to use these relationships to predict the outcome. Therefore, one might believe that the more variables in the model related to the outcome, the greater the model's predictive power. However, based on the principle of parsimony, simple models with fewer predictor variables are preferred over models with many variables (35) for several reasons. First, sparsity, a measure of how small or scanty a model is in terms of variables, is often used to measure interpretability mainly because humans' ability to handle several cognitive entities at a time is limited.

Therefore sparsity makes it easier to check and reason about counterfactuals as can be seen in the Bayesian network part chapters 8 and 9.

Secondly, when too many variables exist in a model, the chances of finding highly correlated predictors (multicollinearity) are almost always high. Including correlated predictors in the same models implies one of the variables is redundant since it adds no new information in explaining the outcome, which weakens the model's predictive power (36). Lastly, too many predictor variables give rise to models which are very complex and dependent on the training data since they try to capture all the patterns/relationships within the data, which makes the model lose its ability to generalize on new data and increases the risk of the model overfitting the training data. Consequently, the model becomes inefficient, inapplicable, incomprehensible (interpretability), and the computational cost of training the model increases significantly.

Therefore, the variable selection process that extracts a subset of appropriate predictor variables from a set of available variables is pivotal in the model development phase, though it is very challenging. With the coming of the Big Data era, which grants access to datasets with dozens, hundreds, or even thousands of predictors, the million-dollar question is, "which subset of these predictors variables will yield the simplest model possible but with high enough predictive power to correctly and efficiently predict instances in the outcome variable?" Prior knowledge from scientific literature was previously the primary raison d'être for the inclusion or exclusion of a variable from a model. This method met its end when literature could not match all research questions asked. This setback gave birth to the development of an endless list of herculean decision rules and algorithms-based variable selection methods (37). However powerful, there is no universal consensus on which method is best under all considerations. Hence, the variable selection method employed should depend on the research priorities, provided the pros and cons of the method are known (38).

The expert-based variable selection technique is one of the most neglected and underused variable selection methods because transportability is seldom a concern. Also, the fear that experts will always only select those variables known to correlate with the outcome could be another reason most researchers shy away from the method since it adds no new information to existing knowledge and can introduce selection bias to the study. However, the importance and impact of experts' contribution to the variable selection process are enormous and will be discussed later.

Collaboration

Tackling the main challenges in the model-building process like those mentioned earlier is not enough. Another less apparent hurdle that hinders machine learning models from attaining their full potential, especially in the healthcare domain, is something that we

like to call "collaboration variety." Outcome prediction models like survival stem from different research groups and labs, with each having little or no data and expertise from another research group because they work in isolation. For example, a tumor recurrence prediction model is sometimes built with clinical, genomics, and even imaging information but will not include parameters like diet, occupation, residence, income, insurance, mode of transportation, and lots of other factors which might influence tumor recurrence. Such models seldom attain optimal predictive performance, and even when they do, they will hardly generalize to other populations. Prediction models in healthcare should not only be limited to healthcare data. As it only includes the patient journey partially.

On the contrary, we propose including patient information beyond admission and discharge with their physical and socio-economic characteristics to have a holistic picture of the patient's journey since all this information contributes to the patient's overall health, as discussed in chapter 3. An excellent example is using marital status, income, and race to predict overall survival in patients with stage III non–small-cell lung cancer (39). However, it might be arduous to develop such a model, especially with the General Data Protection Regulation (GDPR) in place. Therefore, one might have to rely on federated learning techniques, which have seen great success over the past years, especially in terms of collaboration and data observation (40).

Importance of interpretability

Researchers are gradually shifting to more complex models like neural networks and spline models to optimize prediction performance. These models contain many (hyper)parameters that can be adapted to give the model the necessary flexibility to fit the data. Although these models have seen great success in the different areas of radiation therapy, they are not the only or ultimate algorithms suitable for solving healthcare problems. In most cases, more complex models will perform better than simpler models. For example, a neural network will undoubtedly outperform a simple regression model for image recognition tasks. However, for clinical data, simpler models will perform almost as well as these complex models with little effort and assistance from domain experts (41, 42), as in chapter 8. In addition, a large amount of data is needed to train these complicated models properly. For example, self-driving cars use Light Detection and Ranging (lidar) to create 3D maps of the car's surroundings by emitting laser beams in all directions, scanning hundreds of thousand points in a second (43). These scans feed the model with additional information to fill the knowledge gaps that the model might be missing. In radiation oncology, the amount of information (scans) available to train these models in a single hospital is too little, and collaborating with other institutions to increase the sample size to build a decent model is also being challenged chiefly due to patient privacy issues, which means these models are liable to overfitting.

Therefore, one could argue that we favor model performance for the wrong reasons because the internal working mechanism of these complex models is hardly known, hence the name "black box" models. This lack of model interpretability might be problematic for end-users because they can be held accountable for decisions made by these models. Generally, people seldom trust what they do not understand. Therefore, a model which predicts without concrete arguments for its decision will likely not be trusted by its users. On the other hand, interpretable models give vivid and transparent reasoning of how each decision was reached and insights into causality within the training data. Therefore, model interpretability is a valued component of prediction models, especially in healthcare, where stakes are much higher, and accountability is an asset.

Regulations are being put in place to legally protect individuals affected by automated algorithmic systems by granting them the right to demand a logical explanation of how the system reached such a decision. As a result, most research institutions are gradually shifting to interpretable models or seeking ways to interpret the workings of these black-box models. Regulations aside, interpretable models are essential since they give users the feeling of control over what is happening, making their inner workings inherently easier to communicate to stakeholders and less technical audiences. In addition, interpretability can ensure the trained model has not inherited any bias from the training dataset. In other words, interpretability would help debug the trained model by drawing attention to the various adversarial perturbations that might influence the predictions or decisions of the model.

As discussed earlier, models might sometimes perform exceptionally well not because they have been adequately trained but because they include some form of noise inherited from the data the model was derived from. That is making the correct prediction for the wrong reasons (chapter 8). An excellent example of the danger of these opaque Local Interpretable Model-agnostic Explanations (LIME) is that a model developed to distinguish between wolves and huskies was not using the anatomical feature of these animals but rather the surroundings. The model relied heavily on the presence of snow to make its decision and would classify an animal as a wolf if there is snow and huskies otherwise.

As evident from the study by Ribeiro et al. 44, such misguided correlations can never be spotted with black-box models, and if these biases are not identified and tackled before model deployment, it could open up a whole world of moral, ethical, and legal problems. Consider for a second that instead of snow, this model was dependent on more sensitive features like gender or ethnicity to make its decisions. Such instances highlight the value of interpretable models since they can effortlessly identify such misguided dependencies within the model (Chapter 8 and 9) to ensure that clinical, ethical, and legal integrity is not violated. In addition, interpretable models could give a clue into the accuracy of the data used to build the model. For example, a decision tree that splits patients into the different marital statuses for a node that contains patients younger than five years.

Although model interpretability is essential, it depends largely on the target audience. Generally, healthcare, specifically radiation oncology, need machine learning to support and improve clinical practice on two fronts: better cancer care and treatment process automation. Therefore in a situation where the model gives an opinion or makes a prediction like which treatment option is best for a patient or if a said patient is going to have a tumor recurrence within a time interval, model interpretability is essential. However, if the model's decision does not directly affect the patient, like auto contouring of CT scans, model interpretability might be less of an issue. In essence, we need not know the internal workings of the model if their purpose is to automate time-intensive manual tasks to save time for the expert to focus on more critical tasks like patient examination since the expert still has the final decision after manual check to discard, correct, or keep the laborious automated solution of the model. Just like self-driving cars are still built with all the parts of a standard car to allow the rider to take control of the car when it starts behaving abnormally.

The value of experts

The phrase "nothing lasts forever" applies even to predictive models since the model's predictive performance deteriorates over time, a concept known as concept or model drift. Changes in population characteristics and treatment options can affect the established relationships between or within predictor variables and the outcome of interest. For example, high carcinoembryonic antigen (CEA) levels in the pre-pandemic era indicated possible tumor recurrence in non-small cell lung cancer patients (45). In the post-pandemic era, a high CEA level could also indicate a poor covid-19 outcome (46). Therefore a pre-pandemic model with such variable(s) might be challenging to use in the post-pandemic era since a high CEA level could mean either of both outcomes. Therefore, it is crucial to be aware of these potential variable changes and their impact on the model's future performance. Domain experts can quickly adapt to these changes, whereas prediction models will fail due to their established underlying relationship between the features in the model (their lack of a view of the world). These challenges stress the importance of involving experts in every project since they can point out when such life changes occur, like introducing a new treatment (protocol) or new tumor staging system, so that we can detect and adequately adapt these changes at the model level. The methods to detect and deal with concept drift are outside this thesis's scope, but retraining the model is a fast and easy way out.

Missing information within a dataset can sometimes be very informative as well. Therefore, it is worthwhile to thoroughly study this missing information for better processing and check if there is any hidden information behind it because there is always a reason for missing values. For example, a patient might have missing information on the surgery type variable. However, this information might be missing because it was forgetfully not filled in

for some reason. Another possibility could be that the patient is in the "watch and wait" treatment option, which might be deduced by inspecting the other variables like tumor size and pathological lymph node status.

Nevertheless, to make such inferences, a solid understanding of the biological system under study is needed, which can be provided only by domain experts who are also the end-users of the prospective model. Therefore, it is pivotal to include end-users at the beginning of every study to rectify missing information problems such as the example mentioned above, which will be wrongly imputed if not rectified. Also, expert knowledge of variables prone to missingness can guide the data collection process to include known correlates of these prone to missingness variables. These correlated secondary variables can then serve as a surrogate for the model's primary variable of interest. However, this substitution will largely depend on the missingness within the primary variable and the outcome of interest.

Based on variable selection, the experts' selection approach ensures that essential variables are not excluded from the model since biological plausibility is pivotal for interpretable models in healthcare. Secondly, expert suggestions such as combining some predictor variables to form one more informative variable could help reduce the number of candidate variables. This process reduces the actual variable selection time and complexity, leading to a less complicated and inevitably interpretable model since the final subset of variables included in the model might be small. For example, in chapter 4 the difference between pre-treatment and post-treatment blood parameters was used instead of the original variables separately. Furthermore, it would seem that models with predictors chosen by experts would perform equally likely compared to models with predictors chosen purely by an algorithm with a reduced chance of a selection bias as in chapter 8. One possible solution to circumvent the "new knowledge phobia of most researchers" is to complement expert knowledge or opinions with information from the data as implemented in Chapter 9.

In general, experts' involvement in the model building process significantly improves the quality of the model and the training data as they can ascertain if specific patient information is accurate, consistent, or an outlier. An example of such a situation is a stage I cancer patient with metastasis. Although both entries are valid under their respective column, their combination is inconsistent and might affect the model. Furthermore, expert involvement ensures that models like Bayesian networks are clinically correct and in sync with clinical processes, as shown in chapter 8 and 9.

Future perspectives

Generally, interpretable machine learning models are considered a limitation to knowledge advancement, especially those built from experts' options. In other words, because they are easy to build, they are less likely to teach or give us new insights into the relationship

between and within the predictors and the outcome. Nonetheless, future studies should concentrate on methods to complement or blend expert opinions and knowledge from the data, as in chapter 9.

Developing interpretable models such as Bayesian networks is time-consuming, primarily because experts are not that versed with what the researcher requires from them, coupled with their busy schedules. Therefore, future studies should involve experts more in model development projects, so they become acquainted with the process and can give qualitative input in future studies. Organizing short data science seminars will also provide experts with additional knowledge and information on the domain. All these endeavors to get experts acquainted with the model-building process, facilitate future projects, and saves a considerable amount of time.

Variables used to predict the outcomes in this thesis are limited and primarily clinical, with most of the sample sizes in hundreds. Future studies can improve this by including multiple sources of patient information such as omics, pathology, hematological, and even socio-economical data. Furthermore, the distributed learning method, which has not been included in this thesis but is currently in development, is an option to increase the sample size of future studies for a more robust interpretable model.

Conclusion

Big Data is a game-changing revolution that can transform any sector regarding operational efficacy, better processes, and automation. However, Big Data in radiation oncology is still in the infancy stage (comparatively) and requires a lot of collaborative effort between institutions and most importantly between data scientists and clinicians to harness its power before fully benefiting from it.

Regulations aside, caregivers and other medical experts still need assistance translating data into knowledge to improve patient care in routine clinical practice. Prediction models should therefore have end-users in mind during development because providing end-users with a model to assist them in making important decisions that could significantly affect lives will face trust issues, especially when end-users do not fully understand the model's inner workings. Furthermore, when they do not trust the model, they are less likely to use it for decision-making, even if it has good predictive performance. In other words, the models should focus on end-user interpretability first before any other criteria or metrics. Additionally, domain experts, who are most likely the end-users, should be involved in every step of the model development process.

Bibliography

[1] Nicoli-Carr Valeria. Why do patients drop out of clinical trials? 2020. URL https://mdgroup.com/blog/why-do-patients-drop-out-of-clinical-trials/.

[2] Weichung Joseph Shih. Problems in dealing with missing data and informative censoring in clinical trials. *Current controlled trials in cardiovascular medicine*, 3(1):1–7, 2002.

[3] Therese D Pigott. A review of methods for missing data. *Educational research and evaluation*, 7(4):353–383, 2001.

[4] James L Peugh and Craig K Enders. Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of educational research*, 74 (4):525–556, 2004.

[5] Kristel JM Janssen, Yvonne Vergouwe, A Rogier T Donders, Frank E Harrell Jr, Qingxia Chen, Diederick E Grobbee, and Karel GM Moons. Dealing with missing predictor values when applying clinical prediction models. *Clinical chemistry*, 55(5):994–1001, 2009.

[6] Richard D Riley, Joie Ensor, Kym IE Snell, Frank E Harrell, Glen P Martin, Johannes B Reitsma, Karel GM Moons, Gary Collins, and Maarten Van Smeden. Calculating the sample size required for developing a clinical prediction model. *Bmj*, 368, 2020.

[7] Joonyoung Yi, Juhyuk Lee, Kwang Joon Kim, Sung Ju Hwang, and Eunho Yang. Why not to use zero imputation? correcting sparsity bias in training neural networks. *arXiv preprint arXiv:1906.00150*, 2019.

[8] Teresa A Myers. Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication methods and measures*, 5(4): 297–310, 2011.

[9] Rebecca R Andridge and Roderick JA Little. A review of hot deck imputation for survey non-response. *International statistical review*, 78(1):40–64, 2010.

[10] Jau-Huei Lin and Peter J Haug. Exploiting missing clinical data in bayesian network modeling for predicting medical problems. *Journal of biomedical informatics*, 41(1):1–14, 2008.

[11] Shinichi Nakagawa. Missing data: mechanisms, methods and messages. *Ecological statistics: Contemporary theory and application*, pages 81–105, 2015.

[12] MW Heymans and Iris Eekhout. Applied missing data analysis with spss and (r) studio. *Heymans and Eekhout: Amsterdam, The Netherlands: 20Available online: https://bookdown.org/mwheymans/bookmi/[accessed 23 May 2020]*, 2019.

[13] Stef Van Buuren. *Flexible imputation of missing data*. CRC press, 2018.

[14] Fulufhelo V Nelwamondo, Shakir Mohamed, and Tshilidzi Marwala. Missing data: A comparison of neural network and expectation maximization techniques. *Current Science*, pages 1514–1521, 2007.

[15] Paul T Von Hippel. New confidence intervals and bias comparisons show that maximum likelihood can beat multiple imputation in small samples. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(3):422–437, 2016.

[16] YT Mustafaa, V Tolpekin, and A Stein. Application of the em-algorithm for bayesian network modelling to improve forest growth estimates. *Procedia environmental sciences*, 7:74–79, 2011.

[17] Serge Romaric Tembo, Sandrine Vaton, Jean-Luc Courant, and Stéphane Gosselin. A tutorial on the em algorithm for bayesian networks: Application to self-diagnosis of gpon-ftth networks. In *2016 International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 369–376. IEEE, 2016.

[18] Alberto Franzin, Francesco Sambo, and Barbara Di Camillo. bnstruct: an r package for bayesian network structure learning in the presence of missing data. *Bioinformatics*, 33 (8):1250–1252, 2017.

[19] Marco Scutari and Robert Ness. bnlearn: Bayesian network structure learning, parameter learning and inference. *R package version*, 3, 2012.

[20] Maytal Saar-Tsechansky and Foster Provost. Handling missing values when applying classification models. 2007.

[21] Ad Feelders. Handling missing data in trees: surrogate splits or statistical imputation? In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 329–334. Springer, 1999.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[23] Maher Maalouf and Theodore B Trafalis. Rare events and imbalanced datasets: an overview. *International Journal of Data Mining, Modelling and Management*, 3(4):375–388, 2011.

[24] Yang Zhao, Zoie Shui-Yee Wong, and Kwok Leung Tsui. A framework of rebalancing imbalanced healthcare data for rare events' classification: a case of look-alike sound-alike mix-up incident detection. *Journal of healthcare engineering*, 2018, 2018.

[25] Luís Torgo, Paula Branco, Rita P Ribeiro, and Bernhard Pfahringer. Resampling strategies for regression. *Expert Systems*, 32(3):465–476, 2015.

[26] Ikram Chaabane, Radhouane Guermazi, and Mohamed Hammami. Enhancing techniques for learning decision trees from imbalanced data. *Advances in Data Analysis and Classification*, pages 1–69, 2019.

[27] Yuchun Tang, Yan-Qing Zhang, Nitesh V Chawla, and Sven Krasser. Svms modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):281–288, 2008.

[28] Robert C Holte, Liane Acker, Bruce W Porter, et al. Concept learning and the problem of small disjuncts. In *IJCAI*, volume 89, pages 813–818. Citeseer, 1989.

[29] Haibo He and Yunqian Ma. Imbalanced learning: foundations, algorithms, and applications. 2013.

[30] Nguyen Thai-Nghe, Zeno Gantner, and Lars Schmidt-Thieme. Cost-sensitive learning methods for imbalanced data. In *The 2010 International joint conference on neural networks (IJCNN)*, pages 1–8. IEEE, 2010.

[31] Charles Elkan. The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*, volume 17, pages 973–978. Lawrence Erlbaum Associates Ltd, 2001.

[32] Victor S Sheng, Charles X Ling, Ailing Ni, and Shichao Zhang. Cost-sensitive test strategies. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 482. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[33] Kai Ming Ting. Inducing cost-sensitive trees via instance weighting. In *European symposium on principles of data mining and knowledge discovery*, pages 139–147. Springer, 1998.

[34] Charles X Ling, Victor S Sheng, and Qiang Yang. Test strategies for cost-sensitive decision trees. *IEEE Transactions on Knowledge and Data Engineering*, 18(8):1055–1067, 2006.

[35] Joachim Vandekerckhove, Dora Matzke, Eric-Jan Wagenmakers, et al. *Model comparison and the principle of parsimony*. eScholarship, University of California, 2014.

[36] Jim Frost. Multicollinearity in regression analysis: problems, detection, and solutions. *Statistics by Jim. Available online also at: https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/[accessed in Jakarta, Indonesia: December 22, 2018]*, 2017.

[37] Stefan Walter and Henning Tiemeier. Variable selection: current practice in epidemiological studies. *European journal of epidemiology*, 24(12):733–736, 2009.

[38] Farideh Bagherzadeh-Khiabani, Azra Ramezankhani, Fereidoun Azizi, Farzad Hadaegh, Ewout W Steyerberg, and Davood Khalili. A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results. *Journal of clinical epidemiology*, 71:76–85, 2016.

[39] M Vyfhuis, Josephine L Feliciano, Søren M Bentzen, Whitney M Burrows, Elizabeth M Nichols, and Mohan Suntharalingam. Marriage predicts for survival in patients with stage iii non–small-cell lung cancer. *The Journal of Community and Supportive Oncology*, 16(5):e194–e201, 2018.

[40] Timo M Deist, Frank JWM Dankers, Priyanka Ojha, M Scott Marshall, Tomas Janssen, Corinne Faivre-Finn, Carlotta Masciocchi, Vincenzo Valentini, Jiazhou Wang, Jiayan Chen, et al. Distributed learning on 20 000+ lung cancer patients–the personal health train. *Radiotherapy and Oncology*, 144:189–200, 2020.

[41] Biche Osong, Carlotta Masciocchi, Andrea Damiani, Inigo Bermejo, Elisa Meldolesi, Giuditta Chiloiro, Maaike Berbee, Seok Ho Lee, Andre Dekker, Vincenzo Valentini, et al. Bayesian network structure for predicting local tumor recurrence in rectal cancer patients treated with neoadjuvant chemoradiation followed by surgery. *Physics and imaging in radiation oncology*, 22:1–7, 2022.

[42] Timo M Deist, Frank JWM Dankers, Gilmer Valdes, Robin Wijsman, I-Chow Hsu, Cary Oberije, Tim Lustberg, Johan van Soest, Frank Hoebers, Arthur Jochems, et al. Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers. *Medical physics*, 45(7):3449–3459, 2018.

[43] Jeff Hecht. Lidar for self-driving cars. *Optics and Photonics News*, 29(1):26–33, 2018.

[44] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[45] M Grunnet and JB Sorensen. Carcinoembryonic antigen (cea) as tumor marker in lung cancer. *Lung cancer*, 76(2):138–143, 2012.

[46] Bangshun He, Aifang Zhong, Qiuyue Wu, Xiong Liu, Jie Lin, Chao Chen, Yiming He, Yanju Guo, Man Zhang, Peiran Zhu, et al. Tumor biomarkers predict clinical outcome of covid-19 patients. *Journal of Infection*, 81(3):452–482, 2020.

Part VII
Appendices

SUMMARY

Introduction

Machine learning models have seen considerable success, from automating conventional workflows for improving operational efficiency and performance to providing fast, personalized, and reliable recommendations. Examples of real-world applications are Netflix movie recommendations, Tesla self-driving cars, and Amazon speech recognition (Alexa). These applications are possible thanks to the (Big) Data generated and willingly donated by end-users in their daily activities. Patients and caregivers need help translating this level of success to address healthcare challenges and assist them in making personalized decisions for optimal outcomes using the available healthcare and patient (Big) Data.

Therefore, this thesis focuses on healthcare Big Data and the value of interpretable machine learning models for outcome prediction in radiation oncology and highlights the benefits of experts' involvement in the model development process. This thesis is partitioned into a theoretical and practical section, with the theoretical section consisting of two literature review chapters that discuss Big Data in radiation oncology and healthcare in general. In contrast, the practical section contains three parts (Regression, Decision tree, and Bayesian network) representing the fitted interpretable machine learning models, with each part also containing two original research chapters. Chapter 1 briefly introduces the contents of subsequent chapters and provides the blueprint of this thesis.

Part: Big Data

Data has become one of the most valuable commodities in recent years, to the point that it is being likened to crude oil. The healthcare domain is very interested, amongst others, in using (Big) Data to improve cancer care. Thus, chapter 2 gives a gentle introduction to Big Data and its main characteristics in healthcare. Chapter 3 then paints a more detailed picture of Big Data characteristics and its different sources within the confines of radiation oncology. Both chapters 2 and 3 discuss the solutions provided by Big Data for healthcare challenges, with examples where Big Data has improved operational efficiency for clinical excellence. Chapter 3 furthers the discussion on domain applications, barriers, and the future of radiation oncology Big Data.

Part: Regression

Due to the mortality rate, cancer patients are predominantly concerned about how long they have to live, more so for patients whose tumor has metastasized to other parts of the body like the spine. However, accurate prediction of complex endpoints like overall survival is challenging, even for an experienced clinician. Chapter 4 looks at the prediction of progression-free survival and overall survival for cervical cancer patients, while Chapter 5 looks at overall survival within a 1, 3, and 6 months time frame for patients with spinal bone metastases. Both chapters use a Cox proportional hazard regression model and stratify

patients into different survival risk groups to assist caregivers with patient management. Chapter 5 goes a step further and translates the model into a nomogram to provide caregivers with a tool for individualized estimates of survival probabilities for patients with spinal bone metastases.

Part: Decision tree

This part involves a tree-based algorithm, one of the most popular machine learning techniques, because of its non-parametric nature and ability to naturally classify observation into various groups. In addition, they are effortlessly understandable by a less technical audience due to their IF-THEN nature, making them valuable as clinical decision aids. Chapter 6 focuses on discriminating before the start of treatment between elderly cancer patients who will comply with their planned radiotherapy treatment and those likely not to comply using a decision tree (Compliance tree). Chapter 7 extends the application of decision trees to time-to-event data. This chapter developed a decision tree to predict overall survival in women treated in the Netherlands with radiotherapy for squamous cell cervical carcinoma FIGO-stage IIB-IVA (Survival tree) and externally validates the tree on a Korean population. Both chapters found age to be associated with the outcomes. One logical explanation of this result is immunosenescence which makes individuals susceptible to numerous diseases and morbidity, leading to a reduced chance of survival and treatment completion. Chapter 7 compared the survival curves of each risk group created from the decision tree splits of the leaf nodes on the external validation data to ascertain that the model is generalizable.

Part: Bayesian network

Prediction models can assist caregivers with their decision-making by estimating an individual's probability of developing the outcome of interest. However, the biological process which leads to a particular outcome consists of complex relationships interdependent over and within time. This complexity poses a significant challenge for statistical analysis since the likelihood of correlated features is almost inevitable. Also, clinical researchers will have difficulty determining whether or where a variable should be included for model development, making domain experts' contributions indispensable in the model-building process. Predictive models which can probabilistically reason under uncertainty such as Bayesian networks are more suitable to model such information. Chapter 8 tackles some of these problems by eliciting multiple experts' opinions from different countries on the interplay between variables to develop a Bayesian network structure capable of predicting tumor recurrence for rectal cancer patients. This chapter also compares an expert-elicited structure with an algorithmic structure. Chapter 9 builds on the knowledge of Chapter 8 to develop a Bayesian network that predicts two-year survival for lung cancer patients from a symbiotic relationship between experts' opinions and algorithmic knowledge of the data. Both chapters highlight the value of including experts in the model-building process to develop clinically valid Bayesian network structures. Chapter 8 shows the need to include

essential variables in a model based on their availability on a timeline of extraction to have a Bayesian network structure whose reasoning aligns with the clinical process. A bird's-eye view of this part shows that structures developed by multiple experts or received input from multiple experts are clinically more plausible than algorithmic structures or structures developed by a single expert.

Chapter 10 discusses some of the challenges encountered during the model-building process with possible options to tackle these challenges and emphasizes the need to include end-users in the model-building process in healthcare. Finally, this chapter discusses the importance of interpretable machine learning models in healthcare with some future directions for research.

SAMENVATTING

Introductie

Het gebruik van Machine Learning heeft geleid tot aanzienlijke successen, van het automatiseren van werkwijzen voor operationele efficiëntie tot het aanbieden van snelle, gepersonaliseerde en betrouwbare aanbevelingen. Praktijkvoorbeelden hiervan zijn film-aanbevelingen op Netflix, de zelf-rijdende auto van Tesla, en stemherkenning zoals Amazon Alexa. Deze toepassingen zijn mogelijk door beschikbaarheid van "Big Data", gedoneerd door eindgebruikers bij het gebruik van technologie in hun dagelijkse activiteiten. Om dit niveau van succes te evenaren binnen de gezondheidszorg moeten patiënten en zorgverleners geholpen worden. Machine Learning kan helpen bij het maken van gepersonaliseerde beslissingen voor behandeling, gebruik makend van de beschikbare (big) data van voorgaande patiënten.

Deze thesis focust op Big Data binnen de gezondheidszorg, en de waarde van interpreteerbare Machine Learning modellen voor het voorspellen van uitkomsten binnen de radiotherapie. Hierbij wordt de meerwaarde van inclusie van klinische expertise binnen het ontwikkelproces onder de aandacht gebracht. Deze thesis is verdeeld in een theoretisch en praktisch deel, waarbij het theoretisch deel twee literatuuronderzoeken bevat naar het gebruik van Big Data in radiotherapie en de gezondheidszorg in het algemeen. Het praktisch deel bestaat uit drie onderdelen (regressie, beslisbomen en Bayesiaanse netwerken) rondom interpreteerbare machine learning modellen, waarbij ieder onderdeel twee onderzoek hoofdstukken bevat. Hoofdstuk 1 bevat een algemene inleiding, en dient als beschrijving bij deze thesis.

Big Data

Data is de laatste jaren een belangrijk (handels)waar geworden, waardoor de analogie met ruwe olie wordt gemaakt. De gezondheidszorg sector is steeds meer geïnteresseerd om (Big) Data te gebruiken om kankerzorg en behandelingen te verbeteren. Hoofdstuk 2 geeft dan ook een korte introductie in Big Data en de karakteristieken hiervan binnen de gezondheidszorg. Hoofdstuk 3 beschrijft een meer gedetailleerd beeld van Big Data, en de verschillende bronnen die gebruikt kunnen worden binnen de radiotherapie. Beide hoofdstukken beschrijven mogelijkheden op basis van Big Data voor uitdagingen in de gezondheidszorg, met voorbeelden waarbij Big Data operationele efficiëntie en kwaliteit heeft bevorderd. Hoofdstuk 3 beschrijft eveneens de toepassingen in de gezondheidszorg, barrières en de toekomst van radiotherapie en Big Data.

Regressie

Een van de zorgen van kankerpatiënten is de overlevingskans en -duur, zeker voor patiënten met metastasen naar andere lichaamsdelen, zoals bijvoorbeeld de ruggengraat. Helaas zijn accurate voorspellingen voor complexe uitkomsten – zoals overleving – een uitdaging, ook

voor een ervaren arts. Hoofdstuk 4 onderzoekt dan ook de voorspelling van progressie-vrije en algemene overleving bij cervix kanker patiënten, waar hoofdstuk 5 de overleving van patiënten met ruggengraat metastasen onderzoekt bij 1-3-6 maanden na behandeling. Beide hoofdstukken gebruiken het "Cox proportional hazards" regressie model waarbij patiënten in verschillende risicogroepen worden gestratificeerd. Deze risico groepering kan door zorgverleners worden gebruikt voor het inschatten van behandeluitkomsten. Hoofdstuk 5 beschrijft dit niet alleen in statistische getallen, maar ook als visueel nomogram voor het berekenen van individuele inschattingen van overleving in de praktijk.

Beslisbomen

Dit onderdeel van de thesis beschrijft het gebruik van "tree-based algorithms", die populair zijn binnen het machine learning domein vanwege de non-parametrische eigenschappen en het vermogen om observaties op natuurlijke wijze in groepen te classificeren. Verder zijn deze algoritmen makkelijk te begrijpen door een minder technisch publiek vanwege de ALS-DAN keuzes en gevolgen onderliggend aan een beslisboom. Dit maakt ze ook waardevol als klinische beslisondersteuning. Hoofdstuk 6 beschrijft de ontwikkeling van een beslisboom die vóór behandeling de kans berekent dat een oudere patiënt de behandeling gaat/kan doorstaan. Hoofdstuk 7 gaat hierbij verder door het toepassen van beslisbomen op "tijd-tot-voorval" data. Dit hoofdstuk beschrijft de ontwikkeling van een beslisboom voor het voorspellen van overleving bij vrouwen behandeld met radiotherapie voor een cervix plaveiselcarinoom met FIGO-stadium IIB-IVA. Deze beslisboom is extern gevalideerd op een Koreaanse populatie. Beide hoofdstukken vonden leeftijd sterk gecorreleerd met de uitkomst. Een logische verklaring hiervoor is immunosenescentie, waarbij iemand meer vatbaar wordt voor mobiliteiten en leidt tot een verminderde kans op overleving en behandeling adherentie. Hoofdstuk 7 vergelijkt hierbij ook de overlevingscurves van iedere risicogroep (gegenereerd uit de eindresultaten van de beslisbomen) om te valideren of het model generaliseerbaar is over de verschillende datasets en landen.

Bayesiaans netwerk

Voorspellende modellen kunnen zorgverleners ondersteunen bij beslissingen op basis van inschattingen van kansen van het individu op een bepaalde uitkomst. De uitdaging hierbij is dat biologische processen complex zijn, en veel onderlinge afhankelijkheden en relaties hebben. Niet alleen op een bepaald punt in tijd, maar ook over tijd. Deze complexiteit zorgt voor een grote uitdaging voor statistische analyses aangezien er een hoge kans is op correlatie tussen gegevens. Verder is domein expertise nodig in het ontwikkelproces om te bepalen of informatie relevant is in een model, en wanneer deze informatie in het klinisch proces beschikbaar komt. Voorspellende modellen die met kans kunnen rekenen, waarbij onzekerheden in acht worden genomen, zijn daarom meer geschikt voor het modeleren van complexe informatie. Een voorbeeld hiervan zijn Bayesiaanse netwerken. Hoofdstuk 8 beschrijft dan ook de kennisvergaring van meerdere experts uit verschillende landen

voor het beschrijven van causale verbanden tussen medische gegevens. Specifiek voor de observationele, behandel, en uitkomst informatie (lokaal recidief) van patiënten met rectum kanker. Deze informatie wordt gestructureerd in een Bayesiaans netwerk. Deze informatie wordt vergeleken met en Bayesiaans netwerk automatisch gegenereerd door een algoritme, en geëvalueerd op basis van prestatie in het voorspellen van lokaal recidief. Hoofdstuk 9 bouwt voort op de kennis uit hoofdstuk 8 voor ontwikkeling van een Bayesiaans netwerk voor het voorspellen van twee jaar overleving voor longkankerpatiënten, op basis van kennis van experts en data. Beide hoofdstukken (8 en 9) laten de meerwaarde zien van klinische expertise en kennis tijdens het bouwen van klinisch valide Bayesiaanse netwerken. Verder laat hoofdstuk 8 zien hoe informatie op een tijdlijn gepresenteerd kan worden, zodat deze overeen komt met het klinisch proces. Een hogere abstractie laat hierbij zien dat de structuur van causale verbanden, opgemaakt door meerdere experts, klinisch meer plausibel is dan een structuur gebouwd door één expert.

Hoofdstuk 10 bediscussieerd een aantal uitdagingen die tijdens modelontwikkeling naar voren zijn gekomen in de verschillende projecten. Hierbij worden ook mogelijke oplossingen aangedragen, waarbij wordt benadrukt dat klinische experts en eindgebruikers betrokken moeten worden in het model-ontwikkelingsproces. Verder beschrijft dit hoofdstuk de noodzaak van interpreteerbare machine learning modellen in de gezondheidszorg met aanbevelingen voor vervolgonderzoek.

RESEARCH IMPACT

Introduction

Radiation oncology is a fertile field for applying revolutionary Big Data techniques for better cancer care. However, with data sharing across institutions restricted by administrative, political, and legal barriers, radiation oncology Big Data can not function at its full potential. The clinical data science (CDS) research group at Maastricht University, from which this thesis stems, stands on a three-legged platform to front these data barriers problems and provide the necessary assistance to caregivers and patients. The first pillar focuses on developing global FAIR data-sharing infrastructures, the second uses state-of-the-art machine learning techniques to build prediction models from these (FAIR) data, and the third applies these prediction models for better cancer care.

Knowledge dissemination

The three CDS pillars mentioned earlier represent the group's main research areas and reflect this thesis. The theoretical section relates to the antecedent of the first pillar, which is the benefits of a functional FAIR data-sharing infrastructure since access to FAIR data stations will lead to more mature radiation oncology (Big) Data. The analysis section fits the second pillar's goal and develops different interpretable machine learning models for outcome prediction in radiation oncology. The discussion and theoretical sections do not entirely align with the third pillar but discuss issues that hamper the application of these developed models for better cancer care, with one of the root causes being the lack of model interpretability by end-users.

When the logical flow of the patterns of the clinical events leading to an outcome or endpoint of interest is captured within a prediction model such as a Bayesian network, it contributes to the clinical understanding and interpretability of the model's output. In chapter 8, the Bayesian network structure was developed to capture the sequential events on how the variables are extracted in the treatment process on a timeline, making the model's reasoning easier to understand. More so, it prevents the structure from having unrealistic dependencies such that a future variable influences a precedent variable at a particular time point, irregularities that even non-experts can identify when the process is explained.

Cultural focus

Generally, most researchers believe there is a trade-off association between the accuracy and interpretability of a machine learning model. However, chapter 8 does not support this theory and proves this belief is nothing but a "myth" since the predictive performance of the expert Bayesian network is almost as good as the complicated algorithm-based model. The problem is that most researchers allocate the same amount of time to build interpretable and complex models, which will disfavor interpretable models' performance. Interpretable models do require a significant amount of time and effort to construct, with a lot of domain

expert(s) involvement.

This thesis tackles some challenges involved in introducing these developed models to improve healthcare. It will seem end-users are more open and accepting of models that automate time-consuming tasks like auto contouring in the clinic rather than a model that supports decision-making, such as predicting the best treatment option for a patient. For automated models, end-users can visualize the model's output and decide whether to use or discard the model's solution based on how much they agree. This is tricky for prediction models because end-users can not do much with the models' output except when the model is interpretable, as then they can base their judgment on the model's reasoning. Therefore, it all boils down to end-users trust in the model. Chapter 10 provides some suggestions to help increase the end-users' trust in these models, one of which is including the end-users in the building process because if they assist in building the model. They understand how the model works then, there is a high probability of them making use of the model.

Clinical focus

Radiation oncology is a data-rich domain and the perfect field to leverage machine learning techniques to unlock hidden potential knowledge that could assist healthcare professionals and benefit the oncology community. Therefore, this thesis focuses mainly on developing models that can serve as decision aids to caregivers for better patient management. The nomogram in chapter 5 and the decision trees of chapter 6 and 7 can be readily printed to serve as a decision support tool after they have been adequately validated.

Cancer is among the top three leading causes of death worldwide, with metastatic tumors responsible for approximately 90% of all these deaths, making the management of this type of cancer a major clinical challenge. The electronic version of the developed nomogram in chapter 5 provides personalized survival plots, which could come in handy during shared decision-making sessions.

Regression methods are the most common machine learning approach well-versed with physicians and their choice model, especially in decision-making. Unlike regression methods, Bayesian networks are better at dealing with uncertainty related to incomplete domain coverage because the variables used as inputs for the model and their relationships are direct representations of real-world features and their interplay, which is different from other models based on purely mathematical constructs like regression models. Bayesian networks' ability to probabilistically reason about any variable in the structure makes them a valuable tool in any field, more so in medical sciences since explainable visual reasoning increases interpretability. This thesis introduces caregivers to Bayesian networks to increase their knowledge and flexibility in using other machine learning models.

Economical focus

This thesis focuses on developing prediction models which end-users can easily understand and apply to improve healthcare. In addition, chapter 4, 5 and 7 have created risk strata to group patients into subgroups of risk based on their clinical and lifestyle characteristics to assist caregivers with their patient management decisions. These risk stratification can help caregivers reduce costs and improve patient care since they will focus greater attention and resources on the high-risk group patients.

Chapter 6 developed a decision tree capable of discriminating before the start of treatment between patients who will complete their planned radiotherapy treatment and those likely to discontinue (compliance tree). This tree has a two-fold benefit for patients likely to discontinue treatment: the treatment cost and unnecessary treatment-related toxicity. For caregivers, the model saves them valuable time and helps them make better treatment decisions for patients. This chapter also provides some suggestions to help boost the compliance rate for radiotherapy treatment.

The Bayesian network in chapter 8 models the clinical process which leads to tumor recurrence, a predominant concern for most cancer survivors. The structure predicts if a rectal cancer patient will develop a tumor recurrence (True or False) within a specific time after treatment. Patients with a low probability of developing a tumor recurrence could be discharged with limited need for regular check-ups or worries from the patients if their tumor will resurface. However, those predicted to develop a tumor recurrence could be triaged to special hospital services, intensive outpatient case management, and early clinical visits post-discharge. Such applications, therefore, allow for early interventions to reduce readmissions for recurrence patients and maximize cost-effectiveness, especially for patients who are unlikely to develop a tumor recurrence.

Technological focus

This thesis did not develop a fully functional technology per se, but two chapters contain some promising technology still under development. The first product is the nomogram in chapter 5 which is transformed into a shiny application[1]. The app provides personalized predicted survival curves for individuals. It also predicts a patient's survival probability and confidence interval around its predictions at any given time point.

The second product is a transformed version of the expert-elicited Bayesian network structure[2] of chapter 8 for the prediction of local tumor recurrence in rectal cancer patients into an interactive user interface model. The probabilistic dependencies between the variables in the structure align with the clinical process, which means one can use the

[1] https://bich.shinyapps.io/SpinalMets/
[2] https://thomas.zakbroek.com/app/network/rectalcancer

structure to reason forward, from causes to consequences, or backward, and deduce the probabilities of different causes given the consequences.

Although these applications are still under development, the first round of external validation of hopefully many to assess their clinical applicability and generalizability is underway. Their shareable link will ease their integration into the current radiotherapy workflow to assist caregivers in making better decisions for better cancer care when fully developed.

ACKNOWLEDGMENT

This dissertation would not have been possible without the help and support of many.

First and foremost, I would like to express my sincere gratitude to my supervisor, professor Andre Dekker, for giving me the opportunity to pursue my research in his group. Dear Andre, I don't know how to thank you enough for your invaluable guidance, support, and encouragement throughout my Ph.D. journey. Your enthusiasm for my research projects and willingness to support me whenever was truly inspiring. You are always available to advise and make each of your numerous students feel special. I still can't fathom how you manage all these things with such a busy schedule.

To my co-supervisors Inigo and Johan I will simply say thank you because I need a complete section if I have to express how grateful I am for working with you both in the past years. Your expertise, patience, and encouragement have been instrumental in helping me to complete my thesis. I am deeply grateful for your willingness to provide me with feedback and advice and your dedication to helping me reach my goals. I am also thankful for your willingness to listen to my ideas and provide the necessary resources to complete my research. I am truly grateful for your guidance and support as a researcher and father as we discussed my research and fatherhood. I am confident I have gained invaluable knowledge and experience from working with both of you.

I want to also thank Leonard for being a valuable teacher to most of us at CDS, especially during my early days as a researcher. Rianne, thank you for all the support and assistance during my infantry research days. I still remember all you thought me during my first conference in Amsterdam. Alberto, thank you for always being willing to discuss and share some statistical ideas with me.

Thanks to Sonia, Ellen, and Audrey for their support with all my administrative procedures.

To Hajar and Matthijs thank you both for accepting to be my paranymphs. I really do appreciate it.

There are also many of my smart colleagues and friends at MAASTRO and Maastricht University with whom I have shared ideas and enjoyed a wonderful working environment together in these past years. Therefore I wish to say a big thank you to all of them as they have contributed to my research in one way or the other.

Thanks to Roberto and Isabel for all the kindness and encouragement.

I also want to thank Lydie (la mère) and Blaise (chief) for all the support and advice these past years.

A special thanks to the Tschipnang family for their immeasurable support and assistance especially Emmanuel and Jinette.

To Tonton Djoko, Tata Françoise, and Mama Anne I don't have words to express my gratitude for all you have done for me in these past years. You have made these past years feel lighter

I am also grateful to my family and siblings for their unconditional love and support especially my "mother" Careen who has greatly influenced my academic career.

I owe a great deal of gratitude to my parents Osong Daniel and Jessey Biseh for their endless love, support, and encouragement for me to pursue my dreams.

To my raison d'etre Kalel and Kemuel. I first want to apologize to you both for the time not spent with you during my research period and I am very grateful for your understanding.

I am grateful to my partner Nouria. You have been lenient and supportive of my research even when I had to work in the evenings, on weekends, on holidays, or assist in a conference for all these years. Without you and your support, I can only imagine how hard these past years would have been. Thank you!

LIST OF PUBLICATIONS

Published

[1] Biche Osong, Inigo Bermejo, Kyu Chan Lee, Seok Ho Lee, Andre Dekker, and Johan van Soest. Prediction of radiotherapy compliance in elderly cancer patients using an internally validated decision tree. *Cancers*, 14(24), 2022.

[2] Biche Osong, Carlotta Masciocchi, Andrea Damiani, Inigo Bermejo, Elisa Meldolesi, Giuditta Chiloiro, Maaike Berbee, Seok Ho Lee, Andre Dekker, Vincenzo Valentini, et al. Bayesian network structure for predicting local tumor recurrence in rectal cancer patients treated with neoadjuvant chemoradiation followed by surgery. *Physics and Imaging in Radiation Oncology*, 22:1–7, 2022.

[3] I Sanli, B Osong, A Dekker, K TerHaag, SMJ van Kuijk, J van Soest, L Wee, and PC Willems. Radiomics biopsy signature for predicting survival in patients with spinal bone metastases (sbms). *Clinical and Translational Radiation Oncology*, 2022.

[4] Hajar Hasannejadasl, Cheryl Roumen, Henk van der Poel, Ben Vanneste, Joep van Roermund, Katja Aben, Petros Kalendralis, Biche Osong, Lambertus Kiemeney, Inge Van Oort, et al. Predicting erectile dysfunction after treatment for localized prostate cancer. *arXiv preprint arXiv:2110.00615*, 2021.

[5] Biche Osong, Ilknur Sanli, Paul C Willems, Leonard Wee, Andre Dekker, Seok Ho Lee, and Johan van Soest. Overall survival nomogram for patients with spinal bone metastases (sbm). *Clinical and translational radiation oncology*, 28:48–53, 2021.

[6] Esther RC Janssen, Biche Osong, Johan van Soest, Andre Dekker, Nico LU van Meeteren, Paul C Willems, and Ilona M Punt. Exploring associations of preoperative physical performance with postoperative outcomes after lumbar spinal fusion: a machine learning approach. *Archives of Physical Medicine and Rehabilitation*, 2021.

[7] Biche Osong, Andre Dekker, and Johan van Soest. Applications of "big data" in radiation oncology. In Jake Van Dyk, editor, *The Modern Technology of Radiation Oncology: A Compendium for Medical Physicists and Radiation Oncologists*, chapter 11, pages 322–330. Medical Physics Publishing, 4555 Helgesen Dr, 2020.

[8] Joanna Kazmierska, Andrew Hope, Emiliano Spezi, Sam Beddar, William H Nailon, Biche Osong, Anshu Ankolekar, Ananya Choudhury, Andre Dekker, Kathrine Røe Redalen, et al. From multisource data to clinical decision aids in radiation oncology: the need for a clinical data science community. *Radiotherapy and Oncology*, 2020.

[9] Youn Ji Kim, Young Saing Kim, Jin Woo Shin, Biche Osong, and Seok Ho Lee. Prediction scoring system based on clinicohematologic parameters for cervical cancer patients undergoing chemoradiation. *International Journal of Gynecologic Cancer*, 30(11), 2020.

[10] Alberto Traverso, Frank JWM Dankers, Biche Osong, Leonard Wee, and Sander MJ van Kuijk. Diving deeper into models. In *Fundamentals of clinical data science*, pages 121–133. Springer, Cham, 2019.

[11] Akuli B Osong, Andre Dekker, and Johan van Soest. Big data for better cancer care. MA Healthcare London, 2019.

Submitted

[1] Biche Osong, Andre Dekker, Leonard Wee, Johan van Soest, Inigo Bermejo. *"Bayesian network structures for predicting two-year survival in patients diagnosed with non-small cell lung cancer."*

[2] Hajar Hasannejadasl, Cheryl Roumen, Henk van de Poel, Ben Vanneste, Joep van Roermund, Katja Aben, Zhen Zhang, Biche Osong, Lambertus Kiemeney, Inge Van Oort, Renee Verwey, Laura Hochstenbach, Esther Bloemen, Andre Dekker, Rianne R.R. Fijten. *"A data-driven approach to predict urinary incontinence in men with localized prostate cancer"*

Preparatory

[1] Biche Osong, Seok Ho Lee, Ludy Lutgens, Kwang Beom Lee, Leonard Wee, Andre Dekker, Johan van Soest. *"Overall survival prediction in FIGO-stage IIB-IVA cervical cancer patients treated with radiotherapy."*

[2] Biche Osong, Seok Ho Lee, Inigo Bermejo, Leonard Wee, Johan van Soest, Andre Dekker, Maaike Berbee. *"Pathological lymph node staging for intermediate-risk rectal cancer patients"*

[3] Biche Osong, Seok Ho Lee, Johan van Soest, Paul C Willems, Andre Dekker, Inigo Bermejo. *"Decision tree-based survival risk classification for lung cancer patients with spinal bone metastases."*

[4] Biche Osong, Hajar Hasannejadasl, Inigo Bermej, Henk van de Poel, Ben Vanneste, Joep van Roermund, Katja Aben, Johan Van Soest, Lambertus Kiemeney, Inge Van Oort, Renee Verwey, Laura Hochstenbach, Esther Bloemen, Andre Dekker, Rianne R.R. Fijten. *"The value of PROMs for predicting erectile dysfunction in prostate cancer patients with Bayesian network."*

Presentations[3]

[1] Poster *"Bayesian network structure for predicting two-year survival inpatients diagnosed with non-small cell lung cancer"* at the Annual European Lung Cancer Congress (ELCC) Meeting, 30 Mar - 02 Apr, Online.

[2] Oral *"A Bayesian network structure for predicting local recurrence in rectal cancer patients"* at the 39^{th} Annual Meeting of the European Society for Radiotherapy and Oncology (ESTRO), 28 Nov - 01 Dec, Online.

[3] Poster *"Pathological lymph node staging for intermediate-risk rectal cancer patients"* at the 39^{th} Annual Meeting of the European Society for Radiotherapy and Oncology (ESTRO), 28 Nov - 01 Dec, Online.

[4] Oral *"Decision tool for radiotherapy compliance in elderly cancer patients"* at the 45^{th} Annual Meeting of the European Radiation Research Society (ERRS), 13 Sep - 17 Sep 2020, Online.

[5] Poster *"Nomogram for predicting overall survival in patients diagnosed with spinal bone metastases"* at the 45^{th} Annual Meeting of the European Radiation Research Society (ERRS), 13 Sep - 17 Sep 2020, Online.

[6] Poster *"Bayesian network structure for pathological lymph node staging in colorectal cancer patients"* at the 41^{st} EORTC-PAMM winter meeting, 13 Feb - 15 Feb 2020 Stockholm, Sweden.

[7] Poster *"A rapid Learning model for cervical cancer"* at ICCR & MCMA 2019 International Conference, 17 Jun - 21 Jun 2019 Montreal, Canada.

[8] Oral *"Bayesian network structure for predicting two-year survival in patients diagnosed with lung cancer"* at ICCR & MCMA 2019 International Conference, 17 Jun - 21 Jun 2019 Montreal, Canada.

[9] Oral *"Concept drift detection in predictive analytics and machine learning"* at Varian Medical Systems, Inc. 2019 Research Partnership Symposium, 13 May-15 May 2019 Chicago, Illinois.

[10] Poster *"Development of a nomogram for predicting overall survival in patients diagnosed with cervical cancer"* at ESTRO 38, 26 April 2019 - 30 April 2019 Milan, Italy.

[3] International conference presentation only

CURRICULUM VITAE

Biche Osong was born on September 12^{th}, 1987, in Buea, Cameroon. In September 2012, he obtained his bachelor's degree in civil engineering from the National Advanced School of Public Works in Yaounde, Cameroon. After graduation, he worked at Tano Constructions before deciding to switch careers and pursue a master's degree in Biostatistics at Hasselt University.

During his master's program, he built his expertise in statistical data analysis and prediction modeling. His master's thesis titled *"Dynamic Bayesian networks for outcome prediction in lung cancer patients,"* focused on comparing the predictive performance of various machine learning algorithms with Bayesian network to predict acute and late side effects of radiotherapy treatment for lung cancer patients. This thesis sparked his interest in the field of radiation oncology and laid the groundwork for his internship completion at Maastro.

His internship at Maastro molded him into a clinical data scientist and paved the way for him to start a Ph.D. research at Maastro in 2018, where he focused on working with healthcare professionals to develop interpretable machine learning models. During his four years period, he attended several international conferences and won the 2^{nd} place poster presentation prize at the 19^{th} International Conference on the Use of Computers in Radiation Therapy and the 2^{nd} International Conference on Monte Carlo Techniques for Medical Applications (ICCR - MCMA), Young Investigator Award (YIA) at the 45^{th} conference of European Radiation Research Society, travel grants from the European Society for Medical Oncology (ESMO) to attend the 2022 European Lung Cancer Congress (ELCC). He also received recognition for reviewing articles and abstracts for journals and conferences.