

# Predictive modeling of treatment outcome in rectal cancer

Citation for published version (APA):

van Stiphout, R. (2014). *Predictive modeling of treatment outcome in rectal cancer*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20140410rs>

## Document status and date:

Published: 01/01/2014

## DOI:

[10.26481/dis.20140410rs](https://doi.org/10.26481/dis.20140410rs)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# **Predictive modeling of treatment outcome in rectal cancer**

**Ruud van Stiphout**

**Cover illustration**

The cover is an adaptation of Leonardo da Vinci's "Vitruvian Man". The numeric and shape patterns illustrate the increasing amount of data that is available of cancer patients, which makes it possible to predict treatment outcome more accurately.

**Production**

Printing: Datawyse | Universitaire Pers Maastricht

ISBN: 978-94-6159-319-1

© Copyright Ruud van Stiphout, Maastricht 2014

# **Predictive modeling of treatment outcome in rectal cancer**

## **Proefschrift**

Ter verkrijging van de graad van doctor aan de Universiteit Maastricht,  
op gezag van de Rector Magnificus Prof. Dr. L.L.G. Soete,  
volgens het besluit van het College van Decanen,  
in het openbaar te verdedigen  
op donderdag 10 april 2014 om 10.00 uur

door

**Rudi Gerrit Petrus Maria van Stiphout**

Geboren op 31 augustus 1982 te 's-Hertogenbosch



**Promotores**

Prof. Dr. P. Lambin

Prof. Dr. Ir. R.L.M. Peeters

**Co-promotor**

Prof. Dr. G. Lammering (MediClin Robert Janker Clinic, Bonn, Germany)

**Beoordelingscommissie**

Prof. Dr. F.C.S. Ramaekers (voorzitter)

Dr. G.L. Beets

Prof. Dr. R.G.H. Beets-Tan

Prof. Dr. F.M. Mottaghy

# Contents

## ***Introduction***

Chapter 1	Introduction and outline of the thesis	7
Chapter 2	Predicting outcomes in radiation oncology — multifactorial decision support systems	15

## ***Tumor response prediction using PETCT-imaging and blood biomarkers***

Chapter 3	Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential PETCT imaging	53
Chapter 4	Nomogram predicting response after chemoradiotherapy in rectal cancer using sequential PETCT imaging: a prospective study with external validation	71
Chapter 5	Blood biomarkers are helpful in the prediction of response to chemoradiation in rectal cancer: a prospective, hypothesis driven study on patients with locally advanced rectal cancer	89

## ***Predicting long-term follow-up outcomes for decision support***

Chapter 6	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	105
Chapter 7	Two-year disease-free and pathological complete response as clinical intermediate endpoint using pooled data of randomized trials for locally advanced rectal cancer	123

## ***Discussion***

Chapter 8	General discussion and future perspectives	141
	List of abbreviations	157
	Summary	161
	Nederlandse samenvatting	167
	Acknowledgements / dankwoord	173
	Curriculum Vitae	181
	List of publications	183



# Chapter 1

Introduction and outline of the thesis

## Rectal cancer

Cancer is a disease causing uncontrolled division of abnormal cells in a certain part of the body. When cells in the human body become old or damaged, they are replaced with new cells. However, DNA (genetic material) of a cell can become damaged or changed, producing mutations that affect normal cell growth and division. As a consequence superfluous cells may form a mass of tissue called a tumor. A tumor can be benign or malignant. Only malignant tumors are called cancerous and they are a threat to the patient's health since they may invade other adjacent organs, are more prone to grow back after treatment (local recurrence) and their cells can spread to other parts of the body through the blood and lymph systems (metastasis).

Malignant tumor growth in and from the inner wall of the distal part of the large intestine is called rectal cancer. Colorectal cancer (cancer of the colon and rectum combined) is in the modern western world the 3rd cancer, both with respect to incidence rate and mortality rate (about 9% of all cancer cases and cancer related mortalities).<sup>1,2</sup> Only prostate and lung cancer for men and breast and lung cancer for women have a higher incidence, and lung cancer heads the list of mortality. In the Netherlands, colorectal cancer is even the 2nd cancer in men for both incidence and mortality and the 2nd in women for incidence.<sup>1</sup> In the year 2011 a total of 21463 patients were alive in the Netherlands with rectal cancer (<http://www.iknl.nl/>). Rectal cancer is more common in men and the incidence rates increased the last 20 years from 21.8 to 25.2 for men and from 13.4 to 14.2 for women (per 100.000 inhabitants standardized with the European standardized rate). Currently, 59% of the rectal cancer patients survive the first 5 years after diagnosis. This survival rate increased by 20% over the last 40 years due to improvements of the treatment, mainly related to changes and standardization of the surgical approach, better imaging and optimal use of chemo- and radiotherapy before and/or after surgery. Factors that affect the prognosis of rectal cancer are among others the degree of penetration of the tumor through the rectal wall, the possible involvement of the tumor into regional lymph nodes and the presence of distant metastasis.<sup>3</sup> The circumferential margin (CRM) after surgery is also an important predictor of local and distant recurrence and patient survival.<sup>4</sup>

In general, two classifications are often distinguished in rectal cancer: non-locally advanced rectal cancer (NLARC) and locally advanced rectal cancer (LARC). There are various clinical definitions for LARC and NLARC depending on the country of issue, but in general NLARC patients present a clear margin between tumor and mesorectal

fascia (MRF) on magnetic resonance (MR) images. LARC on the other hand is referring to the status in which the tumor invaded the MRF or when the tumor is involved into regional lymph nodes. In other countries, T<sub>3</sub> and/or N<sub>1</sub> is a common definition for LARC (60-70% of rectal cancer patients), while in the Netherlands one of the definitions is positive margins and/or N<sub>2</sub> status (about 15-20% of the tumors). Generally, LARC patients have a worse prognosis with 5-year survival rates of 28% to 59% than NLARC patients with 5-year survival rates of 67% to 74% after treatment.<sup>5</sup> In this thesis only LARC patients have been studied.

## Treatment of rectal cancer

Surgery is the major component of the treatment for rectal cancer patients. However, until 10 years ago, without any other treatment, conventional surgery alone resulted in a high probability of local recurrence.<sup>6</sup> Another surgical technique was introduced and standardized over the last 10 years called the total mesorectal excision (TME). Compared to conventional surgery, this technique resulted in a significant decrease in the number of local recurrences and consequently improved disease free survival.<sup>7,8</sup> The efficacy of TME is however also found to be highly dependent on the specific experience of the surgeon with the technique.<sup>9,10</sup> Radiotherapy (RT) is a treatment modality that can be administered before surgery for rectal cancer to decrease the local recurrence rate. Patients undergoing RT receive ionizing radiation generated by a linear accelerator. The treatment beams are directed and shaped by dedicated treatment planning software in such a way that the tumor receives sufficient dose, while the dose to surrounding healthy tissue and organs is limited. Certain chemotherapy agents help sensitizing radiation in the neoadjuvant setting so that it works better. In order to kill rapid dividing cells, antineoplastic drugs are administered to the patient. These drugs are designed to impair the mitotic part of the cell cycle, required for cell division. The downside is that chemotherapy also affects the cell cycle of healthy rapidly dividing cells like cells in bone marrow, digestive tracts and hair follicles.

In the Netherlands, patients diagnosed with NLARC are treated with either TME surgery alone or surgery after a RT scheme of 5 fractions of 5 Gy (Gray, units or RT dose). This short RT scheme was proven to further significantly reduce probabilities of local recurrence compared to TME alone.<sup>11</sup> The group of patients diagnosed with LARC follows a protocol of long-course scheme of radiotherapy in 28 fractions of 1.8 Gy combined with the chemotherapeutic agent 5-fluorouracil (5FU). This agent is in these

cases administered as capecitabine (Xeloda), which is enzymatically converted into 5FU. In Europe, variations in this long-course treatment schemes are present, for example the RT administration of 45 Gy with a 10 Gy boost and/or a combination of the chemotherapeutic agents capecitabine (a prodrug of 5FU) and oxaliplatin. However, in general it was found that a long-course chemoradiotherapy (CRT) scheme results in significant downstaging and downsizing of the primary tumor.<sup>8,12</sup> In 15-30% of the LARC patients treated with CRT, the response to treatment is excellent, resulting in a complete response confirmed by pathology (pCR). For those patients a less invasive surgical approach or even a watchful waiting approach can be optional to avoid complications and/or a decrease in quality of life due to sphincter loss. In contrast, patients who are the least sensitive to the CRT might benefit from treatment intensification. Therefore it would be useful to identify groups of patients who respond differently to treatment as early as possible, to be able to adapt treatment in an early phase. This is an approach towards individualized medicine for rectal cancer, which will be the central theme in this thesis.

## Individualized treatment and decision tools

In order to identify different risk groups accurately it is important to consider all the information available on a patient. Currently, treatment decisions are based on general guidelines and implicit experience, often only taking into account tumor stage and physical condition of the patient (<http://oncoline.nl>). This limited assessment leads to over-treatment for some patients and inadequate therapy for others, which results not only in a burden to the patient but also to healthcare expenses. However, the amount of available medical information has expanded rapidly over the last few years and will continue to increase due to the development of new (diagnostic) tools like genomic profiling and more advanced imaging techniques. Because of this information expansion, the opportunities for the physician to make a detailed assessment of risk and benefits associated with a specific combination of tumor, patient and treatment characteristics are limited. Computers, and more specifically dedicated decision tools, are better able to deal with these large amounts of information. Therefore, these tools are inevitable in the future with the aim to assist in treatment decision making for cancer.

For locally advanced rectal cancer the treatment decisions will most likely be based on the probability of a patients' response to CRT and the probability that the complete treatment cycle will reduce recurrence rates and therefore increase overall survival in

the long term. To assess individual response in rectal cancer, positron emission tomography (PET) is an important imaging technique. Making a PET scan involves injection of a positron emitting radionuclide into the bloodstream of the patient and the measurement of photons resulting from the interactions between these positrons and electrons within the patients' body. The most commonly used radionuclide in oncology is 18F-fluorodeoxyglucose (FDG). This glucose analog is heavily consumed by brain, kidney and cancer cells. The PET scan reveals functional information and is usually made in combination with a computed tomography (CT) scan for the anatomical information. Many studies have been performed assessing treatment response using PETCT imaging, and it was shown that accurate predictions can be made, especially when assessing PET pre-treatment and during/after the treatment.<sup>13</sup> However, these studies rarely combined PETCT-imaging with clinical factors and they lack external validation. Previously, also blood biomarkers such as carcinoembryonic antigen (CEA), osteopontin (OPN) and interleukin-6 have been studied for rectal cancer and found to be prognostic for outcome.<sup>14-16</sup> A holistic approach in which variables from different sources are combined is expected to generate the most accurate predictions for outcome, as for example has been shown for lung cancer, combining clinical variables with blood biomarkers to predict survival.<sup>17</sup> If in rectal cancer accurate predictions for tumor response can be made by for example using clinical, imaging and biomarker information, treatment decisions can be guided with those predictions. For excellent responders there is an option to avoid surgery to maintain the quality of life for the patient. This decision requires a strong indication that tumor disappeared after treatment, which is mainly assessed using endoscopy. Excellent results have been reported for this wait-and-see approach.<sup>18</sup> In the context of this decision, the prediction models can assist to administer for example a radiotherapy boost when it is predicted that the patient will only respond moderately to CRT. Predicting outcomes that occur later, such as local recurrences, distant metastases and survival, can also be very useful to tailor treatment, especially because these are often the primary outcomes of clinical trials. However, these predictions are hard to make because of the noisy nature of these outcomes, resulting in the requirement that the prediction data have to hold large numbers of patients.

## **Objectives and outline of the thesis**

The main aim of this thesis is to study the prognostic and predictive value of clinical factors, imaging and blood biomarkers for pathologic complete response and follow-up outcome related to recurrences and survival of patients with locally advanced rectal

cancer. The focus of this work is on the development of multivariate prediction models which are able provide a prediction score for every individual patient. To provide sufficient confidence for these predictions all these models are based on large databases with high quality data, externally validated and based on state-of-the-art statistical models. Accurate prediction of response and follow-up outcome would allow treatment modifications in an early stage for certain risk groups, which is expected to reduce toxicity and mortality for these patients. This work is a first step towards computer assisted decision making for rectal cancer.

In **Chapter 2** a general overview of predicting outcome in radiotherapy is given. This involves an introductory section on methodological aspects of outcome prediction and a literature overview of published examples of clinical, treatment, imaging and molecular factors related to outcome. Furthermore, it is described how prediction models should be interpreted and how they can be used as a part of decision making tools in clinical practice. The following chapters are divided into two parts, related to the outcome of subject in rectal cancer. These chapters are put in context of the treatment schedule of LARC in Figure 1.1.

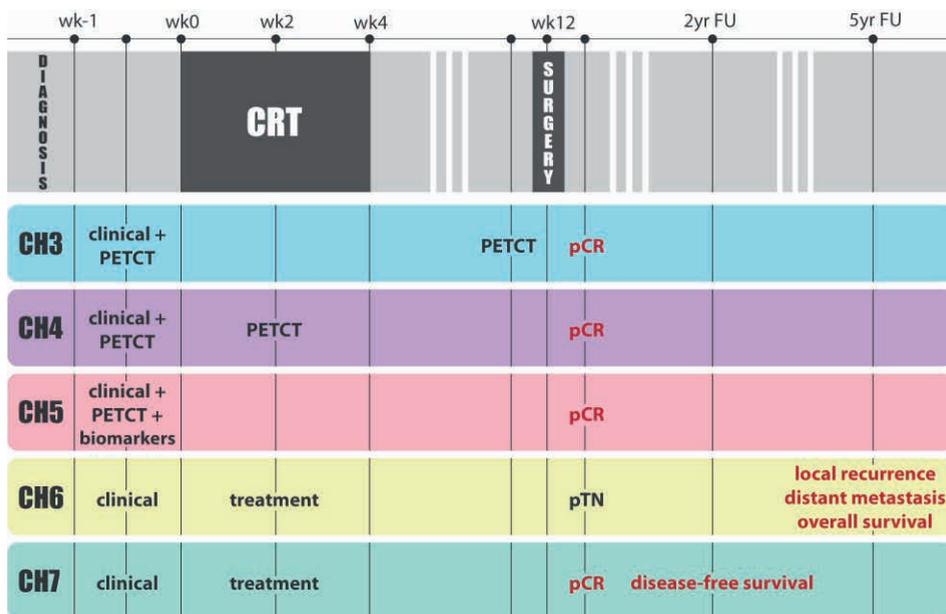
### ***Part 1: Tumor response prediction using PETCT-imaging and biomarkers***

The first part focusses on the prediction of pathological complete response in locally advanced rectal cancer patients. In **Chapter 3**, an externally validated prediction model is developed for pCR using clinical factors and sequential PETCT imaging just before radiochemotherapy and just before surgery. This would allow assistance in making the decision for a wait-and-see policy or reduced surgery for patients with a very high probability of complete response. With the same motivation pCR is also predicted in the prospective sequential PETCT imaging study described in **Chapter 4** but in this case the second PETCT scan is already made after two weeks of radiochemotherapy. This early prediction with the developed model would allow for treatment modifications like radiotherapy boosts or chemotherapy intensification, to increase the number of responding patients. **Chapter 5** describes a prospective study showing that blood biomarkers have an added value to clinical and imaging factors when predicting response.

### ***Part 2: Predicting long-term follow-up outcomes for decision support***

This part involves studies which focus on outcomes recorded in the follow-up time after surgery. These outcomes are important because they are considered as measures for effectiveness of the treatment in elongating patients' lives, reducing the number of

recurrences and improving the quality of life for patients. Knowing which set of combined factors affect these outcomes allows decision making for additional treatment after surgery. **Chapter 6** describes the development of prediction models for the risk of local and distant recurrences and the probability of survival for LARC based on a large clinical dataset containing pooled data from five European randomized clinical trials. This same cohort of patients was used in **Chapter 7** to show the benefits of being two years free of disease and if this is of value to tailor treatment when combined with the knowledge of being a complete responder just after treatment.



**Figure 1.1** Representation of the relationship between the thesis chapters and the treatment scheme of LARC patients. Red data represent evaluated or predicted outcomes. CH: chapter, FU: follow-up, pCR: pathological complete response, PETCT: positron emission tomography combined with computed tomography, pTN: pathological tumor and nodal staging, CRT: chemoradiotherapy, wk: weeks, yr: years

**Chapter 8** includes a general discussion on the developed prediction models and places them into perspective with other related studies. An outlook is provided on how these decision tools may be implemented into daily clinical practice.

## References

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013; 49:1374-403
2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin* 2013; 63:11-30
3. Signaleringscommissie kanker, Kanker in Nederland: trends, prognoses en implicaties voor zorgvraag, 2004, Drukkerij van den Boogaard, Oisterwijk B.V.
4. Hermanek P, Junginger T. The circumferential resection margin in rectal carcinoma surgery. *Tech Coloproctol* 2005; 9:193-9; discussion 199-200
5. Edge SB, Byrd DR, Compton CC, et al: *AJCC Cancer Staging Manual* (ed 7), 2010, Springer, New York
6. Kapiteijn E, Marijnen CA, Colenbrander AC, et al. Local recurrence in patients with rectal cancer diagnosed between 1988 and 1992: a population-based study in the west Netherlands. *Eur J Surg Oncol* 1998; 24:528-35
7. Kapiteijn E, van de Velde CJ. The role of total mesorectal excision in the management of rectal cancer. *Surg Clin North Am* 2002; 82:995-1007
8. Peeters KC, Kapiteijn E, van de Velde CJ. Managing rectal cancer: the Dutch experience. *Colorectal Dis* 2003; 5:423-6
9. Kapiteijn E, Putter H, van de Velde CJ. Impact of the introduction and training of total mesorectal excision on recurrence and survival in rectal cancer in The Netherlands. *Br J Surg* 2002; 89:1142-9
10. Kapiteijn E, van de Velde CJ. Developments and quality assurance in rectal cancer surgery. *Eur J Cancer* 2002; 38:919-36
11. Kapiteijn E, Marijnen CA, Nagtegaal ID, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *N Engl J Med* 2001; 345:638-46
12. Valentini V, Coco C, Cellini N, et al. Ten years of preoperative chemoradiation for extraperitoneal T3 rectal cancer: acute toxicity, tumor response, and sphincter preservation in three consecutive studies. *Int J Radiat Oncol Biol Phys* 2001; 51:371-83
13. Zhang C, Tong J, Sun X, et al. 18F-FDG-PET evaluation of treatment response to neo-adjuvant therapy in patients with locally advanced rectal cancer: a meta-analysis. *Int J Cancer* 2012; 131:2604-11
14. Debuquoy A, Goethals L, Geboes K, et al. Molecular responses of rectal cancer to preoperative chemoradiation. *Radiother Oncol* 2006; 80:172-7
15. Park JW, Lim SB, Kim DY, et al. Carcinoembryonic antigen as a predictor of pathologic response and a prognostic factor in locally advanced rectal cancer patients treated with preoperative chemoradiotherapy and surgery. *Int J Radiat Oncol Biol Phys* 2009; 74:810-7
16. Yoon SM, Kim DY, Kim TH, et al. Clinical parameters predicting pathologic tumor response after preoperative chemoradiotherapy for rectal cancer. *Int J Radiat Oncol Biol Phys* 2007; 69:1167-72
17. Dehing-Oberije C, Aerts H, Yu S, et al. Development and validation of a prognostic model using blood biomarker information for prediction of survival of non-small-cell lung cancer patients treated with combined chemotherapy and radiation or radiotherapy alone (NCT00181519, NCT00573040, and NCT00572325). *Int J Radiat Oncol Biol Phys* 2011; 81:360-8
18. Maas M, Beets-Tan RG, Lambregts DM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011; 29:4633-40

# Chapter 2

## Predicting outcomes in radiation oncology —multifactorial decision support systems

Ruud van Stiphout\*, Philippe Lambin\*, Maud Starmans, Emmanuel Rios-Velazquez, Georgi Nalbantov, Hugo Aerts, Erik Roelofs, Wouter van Elmpt, Paul Boutros, Pierluigi Granone, Vincenzo Valentini, Adrian Begg, Dirk De Ruyscher and Andre Dekker

Published in: Nature Reviews Clinical Oncology 2013; 10(1):27-40

## Abstract

With the emergence of individualized medicine and the increasing amount and complexity of available medical data, a growing need exists for the development of clinical decision support systems based on prediction models of treatment outcome. In radiation oncology, these models combine both predictive and prognostic data factors from clinical, imaging, molecular and other sources to achieve the highest accuracy to predict tumor response and follow-up event rates. In this Review, we provide an overview of the factors that are correlated with outcome—including survival, recurrence patterns and toxicity—in radiation oncology and discuss the methodology behind the development of prediction models, which is a multistage process. Even after initial development and clinical introduction, a truly useful predictive model will be continuously re-evaluated on different patient datasets from different regions to ensure its population-specific strength. In the future, validated decision support systems will be fully integrated in the clinic, with data and knowledge being shared in a standardized, instant and global manner.

## Introduction

Over the past decade, we have witnessed advances in cancer care, with many new diagnostic methods and treatment modalities becoming available,<sup>1</sup> including advances in radiation oncology.<sup>2</sup> The abundance of new options and the progress in individualized medicine has, however, created new challenges. For example, achieving level I evidence is increasingly difficult given the numerous disease and patient parameters that have been discovered, resulting in an ever-diminishing number of 'homogeneous' patients.<sup>3</sup> This reality contrasts to a certain extent with classical evidence-based medicine, whereby randomized trials are designed for large populations of patients. Thus, new strategies are needed to find evidence for subpopulations on the basis of patient and disease characteristics.<sup>4</sup>

For each patient, the clinician needs to consider state-of-the-art imaging, blood tests, new drugs, improved modalities for radiotherapy planning and, in the near future, genomic data. Medical decisions must also consider quality of life, patient preferences and, in many healthcare systems, cost efficiency. This combination of factors renders clinical decision making a dauntingly complex, and perhaps inhuman, task because human cognitive capacity is limited to approximately five factors per decision.<sup>3</sup> Furthermore, dramatic genetic,<sup>5</sup> transcriptomic,<sup>6</sup> histological<sup>7</sup> and microenvironmental<sup>8</sup> heterogeneity exists within individual tumors, and even greater heterogeneity exists between patients.<sup>9</sup> Despite these complexities, individualized cancer treatment is inevitable. Indeed, intratumoral and intertumoral variability might be leveraged advantageously to maximize the therapeutic index by increasing the effects of radiotherapy on the tumor and decreasing those effects on normal tissues.<sup>10-12</sup> The central challenge, however, is how to integrate diverse, multimodal information (clinical, imaging and molecular data) in a quantitative manner to provide specific clinical predictions that accurately and robustly estimate patient outcomes as a function of the possible decisions. Currently, many prediction models are being published that consider factors related to disease and treatment, but without standardized assessments of their robustness, reproducibility or clinical utility.<sup>13</sup> Consequently, these prediction models might not be suitable for clinical decision support systems for routine care.

In this Review, we highlight prognostic and predictive models in radiation oncology, with a focus on the methodological aspects of prediction model development. Some characteristic prognostic and predictive factors and their challenges are discussed in relation to clinical, treatment, imaging and molecular factors. We also enumerate the

steps that will be required to present these models to clinical professionals and to integrate them into clinical decision support systems (CDSS).

## **Methodological aspects**

### ***Factors for prediction***

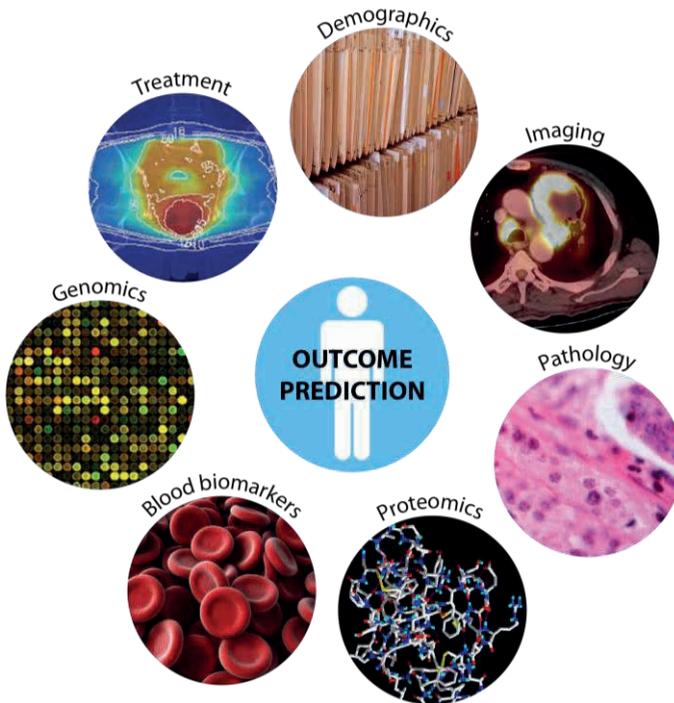
The overall aim of developing a prediction model for a CDSS is to find a combination of factors that accurately anticipate an individual patient's outcome (Figure 2.1).<sup>14</sup> These factors include, but are not limited to, patient demographics as well the results of imaging, pathology, proteomic and genomic testing, the presence of key biomarkers and, crucially, the treatment undertaken. 'Outcome' can be defined as tumor response to radiotherapy, toxicity evolution during follow up, rates of local recurrence, evolution to metastatic disease, survival or a combination of these end points. Although predictive factors (that is, factors that influence the response to a specific treatment) are necessary for decision support, prognostic factors (that is, factors that influence response in the absence of treatment)<sup>15</sup> are equally important in revealing the complex relationship with outcome. Herein, we refer to both of these terms generically as 'features' because, for a predictive model, correlation with outcome must be demonstrable.

### ***Model development stages***

The procedure for finding a combination of features correlated with outcome is analogous to the development of biomarker assays.<sup>16</sup> In that framework, we can distinguish qualification and validation. Qualification demonstrates that the data are indicative or predictive of an end point, whereas validation is a formalized process used to demonstrate that a combination of features is both reliable and suitable for the intended purpose. That is, we need to identify features, test whether they are predictive in independent datasets and then determine whether treatment decisions made using these features improve outcome. The complete cycle of model development entails several stages (Figure 2.2).

In the hypothesis-generation stage, one must consider the end point to predict, the timing of the treatment decision and the available data at these time points. In the data-selection step, a review of potential features is first conducted, ideally by an expert panel. A practical inventory of the available data and sample size calculations are recommended, especially for the validation phase.<sup>17,18</sup> Data from both clinical trials

(high quality, low quantity, controlled, biased selection) and clinical practice (low quality, high quantity, unbiased selection) are useful, but selection biases must be identified in both cases and the inclusion criteria should be equivalent. For all features, including the characteristics of the treatment decision, data heterogeneity is a requirement to identify predictive features and to have the freedom to tailor treatment.



**Figure 2.1** Several data sources for the prediction of outcome after cancer treatment. It is believed that a combination of all the available data with a high variety of sources will predict outcome more accurately than data from individual sources.

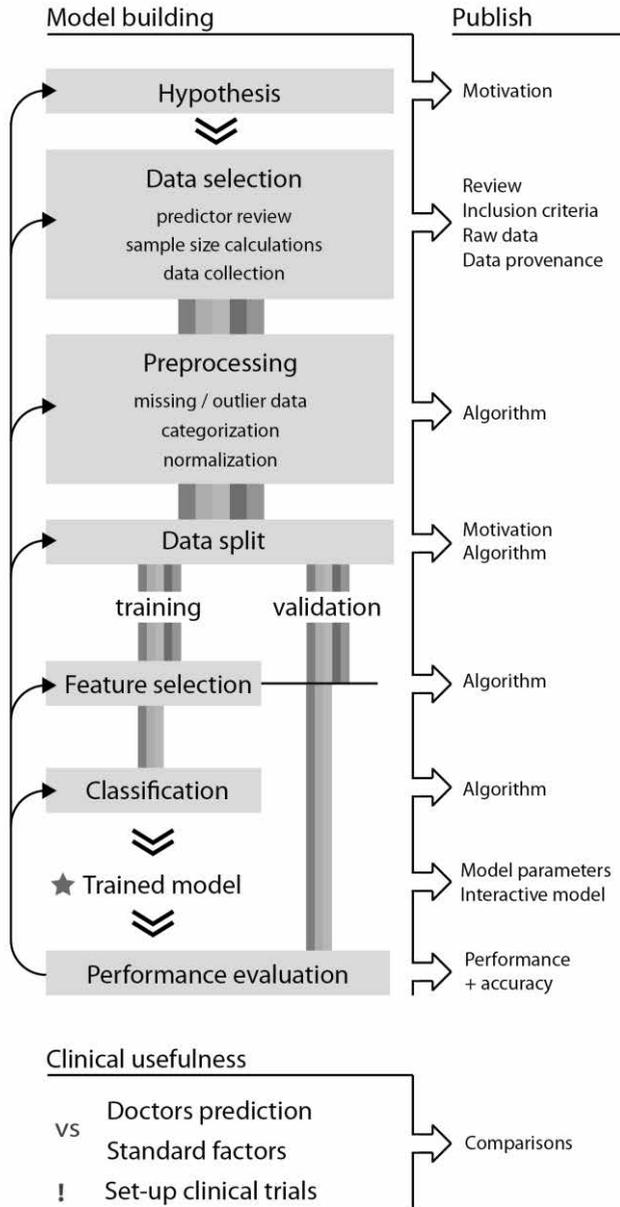
Next, performance measures for models are determined, and include the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity and c-index of censored data.<sup>19</sup> AUC, which has values between 0 and 1 (with 1 denoting the best model and 0.5 randomness), is the most commonly used performance measure. However, for time-to-event models the c-index and hazard ratio are more appropriate because both can handle censored data. The preprocessing stage deals with missing data (imputation strategies; that is, replacing missing values by calculated estimates),<sup>20</sup> identifying incorrectly measured or entered data<sup>21</sup> as well as discretizing (if applicable) and normalizing data to avoid sensitivity for different orders of data

scales.<sup>22</sup> If an external, independent dataset is not available for validation, the available data must be split (in a separate stage) into a model training dataset and a validation set, the latter of which is used subsequently in the validation step. In the feature selection stage, the ratio of the number of evaluated features to number of outcome events must be kept as low as possible to avoid overfitting. When a model is overfitted, it is specifically and exclusively trained for the training data (including its data noise) and, as a result, performs poorly on new data. Data-driven preselection of features is, therefore, recommended.<sup>23</sup> Univariate analyses are commonly used to prioritize the features—that is, testing each feature individually and ranking them on their strength of correlation with outcome.

### ***Predicting outcomes***

In the next stage, the input data are fed into a model that can classify all possible patient outcomes. Traditional statistical<sup>24</sup> and machine-learning models<sup>25</sup> can be considered. For two or more classes (for example, response versus no response), one might consider logistic regression, support vector machines, decision trees, Bayesian networks or Naive Bayes algorithms.<sup>26,27</sup> For time-to-event outcomes, whether censored or not, Cox proportional hazards models<sup>28</sup> or the Fine and Gray model<sup>28</sup> of competing risks are most common. The choice of model depends on the type of outcome (for example, logistic regression for two or more outcomes, or Cox regression for survival-type data) and the type of input data (for example, Bayesian networks require categorized data, whereas support vector machines can easily deal with continuous data). In general, several models with similar properties can be tested to find the optimal model for the available data. A simple model is, however, preferred because it is expected to be robust to a wider range of data than a more complex model.

Performance on the training dataset is upwards-biased because the features were selected. Thus, external validation data must be used, which can be derived from a separate institute or independent trial. When data are limited, internal validation can be considered using random split, temporal split or k-fold cross-validation techniques.<sup>29</sup> The developed model should have a benefit over standard decision making, and must be assessed prospectively in the clinic in the penultimate stage of development. Models must be compared against predictions by clinicians<sup>30,31</sup> and to standard prognostic and predictive factors.<sup>32</sup> Critically, to demonstrate the improvement of patient outcome, quality of life and/or reduced toxicity,<sup>33</sup> clinical trials must be conducted whereby the random assignment of patients is based on the prediction model output.



**Figure 2.2** Schematic overview of methodological processes in decision support system development, describing model development, assessment of clinical usefulness and what ideally to publish. The colored, parallel lines represent heterogeneous data, which have been split early for independent validation (but without internal cross-validation).

Fulfilling this requirement will generate the final evidence that the model is improving healthcare by comparing, in a controlled way, the tailored treatments with standard treatments in the clinic.

Finally, the prediction models and data can be published, enabling the wider oncological community to evaluate them. Full transparency on the data and methodology is the key towards global implementation of the model into CDSSs. This suggestion is similar to clinical ‘omics’ publications for which the raw data, the code used to derive the results from the raw data, evidence for data provenance (the process that led to a piece of data) and a written description of nonscriptable analysis steps are routinely made available.<sup>34</sup> In practice, this cycle of development usually begins by identifying clinical parameters, because these are widely and instantly available in patient information systems and clinical trials. These clinical variables also form the basis for extending prediction models with imaging or molecular data.

## Clinical features

Decision making in radiotherapy is mainly based on clinical features, such as the patient performance status, organ function and grade and extent of the tumor (for example, as defined by the TNM system). In almost all studies, such features have been found to be prognostic for survival and development of toxicity.<sup>35-37</sup> Consequently, these features should be evaluated in building robust and clinically acceptable radiotherapy prognostic and predictive models. Moreover, measurement of some clinical variables, such as performance status, can be captured with minimal effort.

Even the simplest questionnaire, however, should be validated as is the case for laboratory measurements of organ function or parameters measured from blood.<sup>38,39</sup> Furthermore, a standardized protocol should be available to ensure that comparisons are possible between centers and questionnaires over time.<sup>40</sup> Moreover, why specific features were chosen for measurement should be clearly explained. For example, if haemoglobin measurements were only taken in patients with fatigue, the resulting bias would demand caution when including and interpreting the measurements. Only when clinical parameters are recorded prospectively with the same scrutiny as laboratory measurements will observational studies become as reliable as randomized trials.<sup>41,42</sup>

Toxicity measurements and scoring should also build on validated scoring systems, such as the Common Terminology Criteria for Adverse Events (CTCAE), which can be scored by the physician or patient.<sup>43,44</sup> Indeed, a meta-analysis showed that high-quality toxicity assessments from observational trials are similar to those of randomized trials.<sup>45,46</sup> However, a prospective protocol must clarify which scoring system was used and how changes in toxicity score were dealt with over time with respect to treatment.

Finally, to ensure a standardized interpretation, the reporting of clinical and toxicity data and their analyses should be performed in line with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement for observational studies and genetic association studies, which is represented as checklists of items that should be addressed in reports to facilitate the critical appraisal and interpretation of these type of studies.<sup>47,48</sup>

## Treatment features

Currently, image-guided radiotherapy (IGRT) is a highly accurate cancer treatment modality in delivering its agent (radiation) to the tumor.<sup>49</sup> Furthermore, very accurate knowledge of the effects of radiation on normal tissue has been obtained.<sup>50</sup> With modern radiotherapy techniques, such as intensity-modulated radiotherapy, volumetric arc therapy or particle beam therapy, the treatment dose can be sculpted around the target volume with dosimetric accuracy of a few percentage points. IGRT ensures millimeter precision to spare the organs at risk as much as possible.<sup>51</sup>

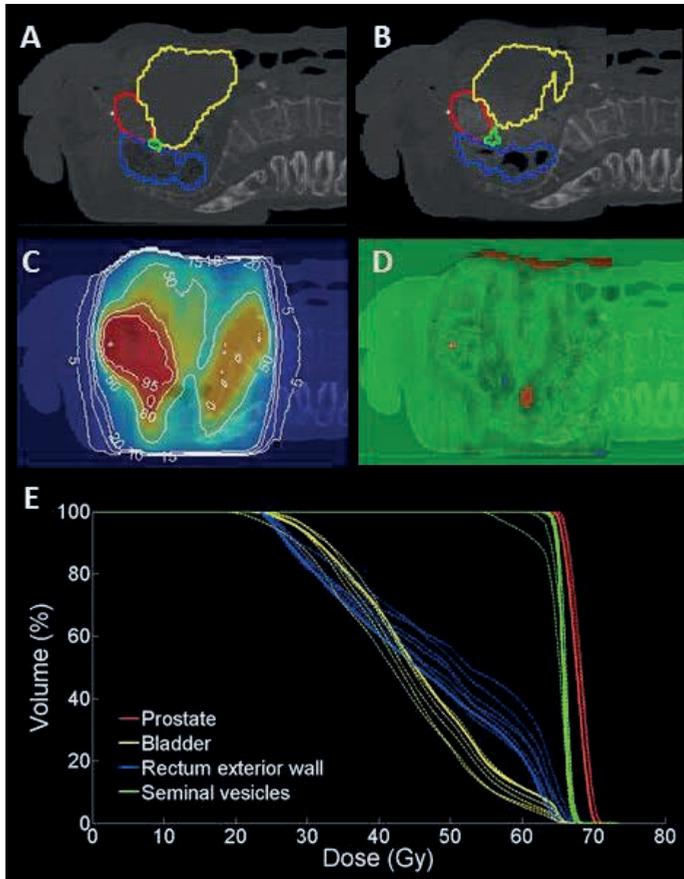
For prediction modeling, recording features that are derived from planned spatial and temporal distribution of the radiotherapy dose is crucial. Additionally, features must be recorded that describe the efforts undertaken during treatment to ensure that the dose is delivered as planned (that is, *in vivo* dosimetry); a delicate balance exists between tumor control and treatment-related toxicity.<sup>52</sup> Additional therapies, such as (concurrent) chemotherapy, targeted agents and surgery, and their features must also be recorded because these have various effects on outcome.<sup>32,53</sup> An example is the difference between concurrent versus sequential chemoradiation, which has a major influence on the occurrence of acute oesophagitis that induces dysphagia.<sup>54</sup>

With respect to the spatial dimension of radiotherapy, how to combine information about the spatially variable dose distribution for every subvolume of the target tumor (or organ) with the global effect to the tumor or adjacent normal tissue remains indeterminate. Dose-response relationships for tumor tissues are often reported in terms of mean (biologically equivalent) dose, although voxel-based measures have also been reported.<sup>55</sup> Mean doses or doses to a prescription point inside the tumor are easily determined and reported and may suffice for many applications. However, spatial characteristics might be more relevant in personalized approaches to ensure radioresistant areas of the tumor receive higher doses.<sup>55</sup> For normal tissue toxicity, dose features including the mean and maximum dosage, as well as the volume of the normal tissue receiving a certain dose, are important. For example, V20 <35% is a common threshold to prevent lung toxicity.<sup>56</sup>

Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small-cell lung cancer was first described in 1991.<sup>57</sup> In 2010, a series of detailed reviews of all frequently irradiated organs (the QUANTEC project) was described,<sup>50</sup> showing that, as for the tumor, care must be taken when assessing dose at the organ level. For example, in some organs, the volume receiving a certain dose is important (such as the oesophagus or lung) because of their proximity to other vital structures, whereas the maximum dose to a small region of other organs might be most important (such as for the spinal cord) because preserving its post-treatment function is crucial. Predicting complications to normal tissue is an active research area in ongoing, large, prospective multicentric projects, including ALLEGRO<sup>58</sup>, and others.<sup>59-61</sup>

Although important, in general one must be careful about relying completely on planned-radiotherapy dose-based predictions because patients display wide variability in toxicity development. The reasons for this variability include many known clinical and molecular-based features as well as the quality of the treatment execution. The focus on the planned radiotherapy dose distribution as the prime determinant of outcome is perhaps the most common pitfall in prediction models because deviations from the original plan during the time of treatment frequently occur.<sup>62</sup> The accuracy of prediction models is expected to increase when measured dose is used, as this measure reflects the effect of radiotherapy most accurately. Figure 2.3 shows an example of these variations in a patient with prostate cancer. Dose reconstructions (2D and 3D), Gamma Index calculations and dose-volume histograms during treatment can help in identifying increasingly accurate dose-related features<sup>63,64</sup> such as radiation pneumonitis<sup>65</sup> and oesophagitis<sup>66</sup>.

The temporal aspect of fractionated radiotherapy is also an active area of research. The fact that higher radiation doses are required to control a tumor when treatment is prolonged is well-known, and increasing evidence suggests that accelerated regimens giving the same physical dose can improve outcome.<sup>67,68</sup>



**Figure 2.3** The importance of considering measured dose for outcome prediction for a patient with prostate cancer. (A) Original planning CT scan that includes contours of the prostate (red), bladder (yellow), exterior wall of the rectum (blue) and seminal vesicles (green). (B) Contoured CT scan after 16 fractions of radiotherapy. (C) Reconstructed 3D dose after 16 fractions of radiotherapy. (D) Calculated dose differences (expressed as a 3D Gamma Index) after 16 fractions of radiotherapy. (E) Dose–volume histograms at fractions 1, 6, 11, 16, 21 and 26 (dashed lines) as well as pre-treatment histograms (solid lines). Clear deviations are visible from the planned dose–volume histogram for the rectum and bladder.

A multicentric analysis of patients with head-and-neck cancer treated with radiotherapy alone showed that the potential doubling time of the tumor before

treatment was not a predictor for local control.<sup>69</sup> Alongside the classic explanation of accelerated repopulation,<sup>70</sup> changes in cell loss, hypoxia and selection of radioresistant stem cells have each been suggested as underlying causes of this observation, the possible implications of which include shorter overall treatment times with higher doses per fraction and the avoidance of breaks during treatment.<sup>71,72</sup> Overall, treatment time is an accessible feature that is correlated with local failure in several tumor sites.<sup>73,74</sup>

Ideally, the spatial and temporal dimensions of radiotherapy would be exploited by showing a fractional dose distribution in a tumor radioresistance (and normal-tissue radiosensitivity) map that is continuously updated during treatment. However, such an image of radioresistance does not yet exist. If it did, CDSSs would guide the planning and modification of the spatial and temporal distribution of radiation in such a way as to maintain or improve the balance between tumor control and the probability of normal tissue complications continuously during treatment, instead of the current approach that delivers radiation as planned with an identical dose to the tumor as a whole.

## Imaging features

Medical imaging has a fundamental role in radiation oncology, particularly for treatment planning and response monitoring.<sup>75,76</sup> Technological advances in noninvasive imaging—including improved temporal and spatial resolution, faster scanners and protocol standardization—have enabled the field to move towards the identification of quantitative noninvasive imaging biomarkers.<sup>77-79</sup>

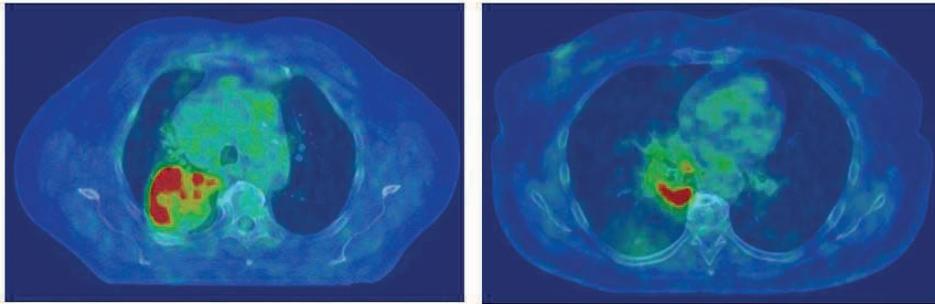
Metrics based on tumor size and volume are the most commonly used image-based predictors of tumor response to therapy and survival,<sup>80-87</sup> and rely on CT and MRI technology for 3D measurement.<sup>88-90</sup> Although used in clinical practice, tumor size and volume measurements are subject to interobserver variability that can be attributed to differences in tumor delineations.<sup>91,92,85-87</sup> Moreover, the optimal measurement technique and definitions of appropriate response criteria, in terms of changes in tumor size, are unclear.<sup>93</sup> Additionally, tumor motion and image artifacts are additional sources of variability.<sup>94,95</sup> To overcome these issues, automated tumor delineation methods have been introduced,<sup>96-99</sup> on the basis of, for example, the selection of ranges of Hounsfield units (which represent the linear attenuation coefficient of the X-ray beam by the tissue) on CT that define a certain tissue type, or calculation of the gradient of an image (mathematical filter) to reveal the borders between tissue types.

Extensive evaluation, however, is needed before these methods can be used routinely in the clinic.<sup>100-102</sup>

A commonly used probe for the metabolic uptake of the tumor is <sup>18</sup>F-fluorodeoxyglucose (FDG) for PET imaging.<sup>103,104</sup> The pre-treatment maximum standardized uptake value (SUV, which is the normalized <sup>18</sup>F-deoxyglucose uptake for an injected dose according to the patient's body weight) is strongly associated with overall survival and tumor recurrence in a range of tumor sites, including the lung, head and neck, rectum, oesophagus and cervix.<sup>105-111</sup> Furthermore, several studies have shown that changes in SUV during and after treatment are early predictors of tumor recurrence.<sup>112-115</sup> FDG-PET measurements, however, are dependent on a number of factors, including injected dose, baseline glucose concentration, FDG clearance, image reconstruction methods used and partial-volume effects.<sup>116,117</sup> Standardization of these factors across institutions is, therefore, fundamental to enable comparisons and validation of data from FDG-PET imaging.<sup>118,119</sup>

Multiple studies have shown that diffusion-weighted MRI parameters, such as the apparent diffusion coefficient (ADC), which is a measure of water mobility in tissues, can accurately predict response and survival in multiple tumor sites.<sup>120-124</sup> However, lack of reproducibility of ADC measurements, due to lack of standardization of instruments between vendors and to lack of internationally accepted calibration protocols, remains a bottleneck in these types of studies.<sup>125</sup> Evaluations of different time points in dynamic contrast-enhanced MRI have also been used to describe tumor perfusion.<sup>90,126-128</sup> Indeed, hypothesis-driven preclinical<sup>129</sup> and xenograft studies support these clinical studies. For example, assessment of the correlation of features from imaging (such as lactate level and the extent of reoxygenation) with tumor control is possible.<sup>130,131</sup>

Increasingly advanced image-based features are currently being investigated. For example, routine clinical imaging can capture both tumor heterogeneity and post-treatment changes, which can be analyzed to identify functional biomarkers (Figure 2.4). Changes in Hounsfield units in contrast-enhanced CT are directly proportional to the quantity of contrast agent present in the tissue and have been used as a surrogate for tumor perfusion.<sup>132,133</sup> Indeed, reductions of Hounsfield units following treatment have been used to evaluate treatment response in rectal, hepatic and pulmonary cancers.<sup>134,135</sup>



**Figure 2.4** Axial 18F-deoxyglucose–PET and CT images of two different patients with non-small-cell lung cancer. Tumor imaging biomarkers describing, for example, textural heterogeneity, FDG uptake and tumor size can be assessed noninvasively before, during and after radiotherapy and associated with treatment outcome. Abbreviations: FDG, 18F-fluorodeoxyglucose; NSCLC, non-small-cell lung cancer.

Standardizing the extraction and quantification of a large number of traits derived from diagnostic imaging are now being considered in new imaging marker approaches.<sup>79</sup> Through advanced image-analysis methods, we can quantify descriptors of tumor heterogeneity (such as variance or entropy of the voxel values) and the relationship of the tumor with adjacent tissues.<sup>136-138</sup> These analytical methods enable high-throughput evaluation of imaging parameters that can be correlated with treatment outcome and, potentially, with biological data. Indeed, qualitative imaging parameters on CT and MRI scans have been used to predict mRNA abundance variation in hepatocellular carcinomas and brain tumors.<sup>139-141</sup> Furthermore, a combination of anatomical, functional and metabolic imaging techniques might be used to capture pathophysiological and morphological tumor characteristics in a noninvasive manner, including apparent intratumoral heterogeneity.<sup>142</sup>

## Molecular features

Biological markers are also valuable clinical decision support features; these include prognostic and predictive factors for outcomes, such as tumor response and normal-tissue tolerance. Despite these strengths, trials of molecular biomarkers are prone to experimental variability; for this reason standardizing assay criteria, trial design and analysis are imperative if multiple molecular markers are to be used in predictive modeling.<sup>16</sup>

## ***Tumor response***

Next to tumor size, tumor control after radiotherapy is largely determined by three criteria: intrinsic radiosensitivity, cell proliferation and the extent of hypoxia.<sup>143</sup> In addition, large tumors intuitively require higher doses of radiation than small tumors because there are simply more cells to kill; this requirement is true even if intrinsic radiosensitivity, hypoxia and repopulation rates are equal. Several approaches have been developed to measure these additional three parameters to predict tumor response to radiotherapy.

### *Intrinsic radiosensitivity*

Malignant tumors display wide variation in intrinsic radiosensitivity, even between tumors of similar origin and histological type.<sup>144</sup> Attempts to assess the radiosensitivity of human tumors have relied on determining the *ex vivo* tumor survival fraction.<sup>145</sup> Those studies and others have shown that tumor cell radiosensitivity is a significant prognostic factor for radiotherapy outcome in both cervical<sup>146</sup> and head-and-neck<sup>147</sup> carcinomas. However, these colony assays suffer from technical disadvantages, that include a low success rate (<70%) for human tumors and the time needed to produce data, which can be up to several weeks.

Other studies have included assessments of chromosome damage, DNA damage, glutathione levels and apoptosis.<sup>148</sup> Indeed, some clinical studies using such assays have shown correlations with radiotherapy outcome, whereas others have not.<sup>149</sup> However, these cell-based functional assays only have limited clinical utility as predictive assays, despite being useful in confirming a mechanism that underlies differences in the response of tumors to radiotherapy. For example, some studies have provided encouraging data showing that immunohistochemical staining for  $\gamma$ -histone H2AX, a marker of DNA damage, might be a useful way to assess intrinsic radiosensitivity very early after the start of treatment.<sup>150,151</sup> Double-stranded breaks are generated when cells are exposed to ionizing radiation or DNA-damaging chemotherapeutic agents, which rapidly results in the phosphorylation of  $\gamma$ -histone H2AX.  $\gamma$ -histone H2AX is the most sensitive marker that can be used to examine the DNA damage and its subsequent repair, and it can be detected by immunoblotting and immunostaining using microscopic or flow cytometric detection. Clinically, two biopsies (one before and one after treatment) are needed to assess the  $\gamma$ -histone H2AX status, which is not always easy to implement in practice.

### *Hypoxia*

Tumor hypoxia is the key factor involved in determining resistance to treatment and malignant progression; it is a negative prognostic factor after treatment with radiotherapy, chemotherapy and surgery.<sup>152,153</sup> Indeed, some data show that hypoxia promotes both angiogenesis and metastasis and, therefore, has a key role in tumor progression.<sup>154</sup> Although a good correlation has been demonstrated between pimonidazole (a chemical probe of hypoxia) staining and outcome after radiotherapy in head-and-neck cancer,<sup>155</sup> the same relationship has not been found in cervical cancer.<sup>156</sup> In light of these contrasting results, one of the hypotheses put forward to explain this is that hypoxia tolerance is more important than hypoxia itself.<sup>157</sup>

The use of fluorinated derivatives of such chemical probes also enables their detection by noninvasive PET.<sup>158-160</sup> Although this approach requires administration of a drug, it does benefit from sampling the whole tumor and not just a small part of it. Another possible surrogate marker of hypoxia is tumor vasculature; the prognostic significance of tumor vascularity has been measured as both intercapillary distance (thought to reflect tumor oxygenation) and microvessel density (the 'hotspot' method that provides a histological assessment of tumor angiogenesis). Some studies have found positive correlations with outcome, mainly using microvessel density in cervical cancer, whereas others have shown negative correlations.<sup>161</sup> Some concerns have been raised about the extent to which biopsies taken randomly truly represent the usually large, heterogeneous tumors.

### *Proliferation*

If the overall radiotherapy treatment time is prolonged, for example, for technical reasons (breakdown of a linear particle accelerator) or because of poor tolerance by the patient to the treatment, higher doses of radiation are required for tumor control, clearly indicating that the influence of tumor proliferation is important.<sup>162</sup> Although proliferation during fractionated radiotherapy is clearly an important factor in determining outcome, reliable measurement methods are not yet available. To understand why radiation leads to an accelerated repopulation response in some tumors and not in others, a greater understanding of the response at both the cellular and molecular level is required.

### ***Normal-tissue tolerance***

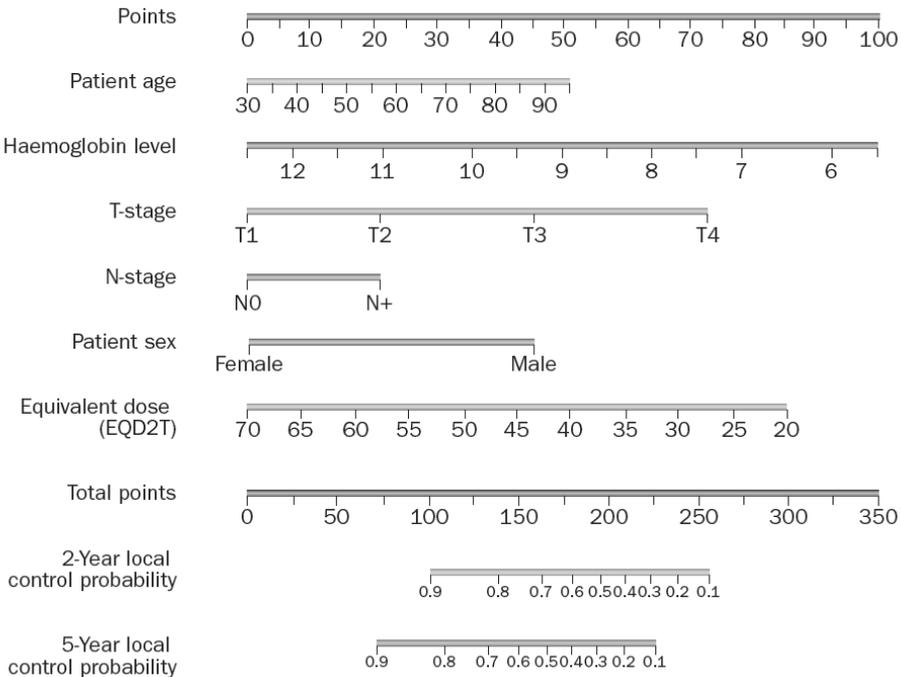
Inherent differences in cellular radiosensitivity among patients dominate normal-tissue reactions more than other contributing factors.<sup>163</sup> That is, the radiation doses given to most patients might in actuality be too low for an optimal cure because 5% of patients are very sensitive; these 5% of patients are so sensitive that they skew what is 'optimal' radiotherapy to the lower end of the spectrum, to the detriment of the majority of patients who are not as sensitive. Future CDSSs should be able to distinguish such overly sensitive patients and classify them separately so they receive different treatments to the less-sensitive patients.

Several small<sup>164</sup> and large<sup>165</sup> *in vitro* studies found a correlation between radiosensitivity and severity of late effects, namely radiation-induced fibrosis of the breast, but these findings were not consistent because no standardized quality assurance exists for radiotherapy *in vivo*.<sup>166,167</sup> Similar discrepancies were later found using rapid assays that measure chromosomal damage,<sup>168</sup> DNA damage<sup>169</sup> and clonogenic cell survival.<sup>170</sup> For example, the lymphocyte apoptosis assay has been used in a prospective trial as a stratification factor to assess late toxicity using letrozole as radiosensitizer in patients with breast cancer.<sup>171</sup> Cytokines such as TGF- $\beta$ , which influences fibroblast proliferation and differentiation are known to have a central role in fibrosis and senescence.<sup>172,173</sup> Currently, the relationships between the lymphocytes predictive assay, TGF- $\beta$  and late complications are purely correlative and a clear molecular explanation is lacking. Genome-wide association studies (GWAS) and the analysis of single nucleotide polymorphisms (SNPs) in candidate genes have also shown promise in identifying normal-tissue tolerance,<sup>174,175</sup> although these do not often validate results from independent studies.<sup>176</sup> In general, the problem with all these studies has been the wide experimental variability rather than interindividual differences in radiosensitivity. Normal tissue tolerance is the dose-limiting factor for the administration of radiotherapy, and therefore any CDSS should be based on predictors of tumor control and the probability of complications.

### **Representation of predictions**

Although the decisions made in the process of developing predictive models will determine the characteristics of a multivariate model (for example, which features are selected and the overall prediction accuracy), the success of the model depends on other factors, such as its availability and interactivity, which increases the

acceptability. Even models based on large patient populations, with proper external validation, can fail to be accepted within the healthcare community if the model and its output are not easily interpretable, if there is a lack of opportunity to apply the model or if the clinical usefulness is not proven or reported.<sup>177</sup>



**Figure 2.5** A published nomogram for local control in patients with cancer of the larynx treated with radiotherapy. Clinical and treatment variables are associated with local control status at follow-up durations of 2 and 5 years. The predictors are age of the patient (in years), haemoglobin level (in mmol/l), clinical tumor stage (T-stage), clinical nodal stage (N-stage), patient's sex and equivalent dose (in Gy). A probability for local control can be calculated by drawing a vertical line from each predictor value to the score scale at the top—'points'. After manually summing up the scores, the 'total points' correspond to the probability of local control, which are estimated by drawing a vertical line from this value to the bottom scales to estimate local control.<sup>180</sup>

Although some models, such as decision trees, implicitly have a visual representation that is somewhat interpretable, most models do not. One highly interpretable representation of a set of features is the nomogram.<sup>178</sup> The nomogram was originally used in the early 20<sup>th</sup> century to make approximate graphical computations of mathematical equations. In medicine, nomograms have experienced a revival, reflected by the increasing number of studies reporting them.<sup>179-183</sup> Figure 2.5 shows

an example of a published clinical nomogram of local control in larynx cancer in which values for the selected features directly relate to a prediction score. The sum of these scores corresponds to a probability of local control within 2 or 5 years.<sup>180</sup>

Another idea for increasing acceptability of computer-assisted personalized medicine is to make prediction models available on the internet. If interactive, peer-reviewed models are provided with sufficient background information, clinicians can test them using their own patient data. Such a system would provide retrospective validation of the multiple features by the wider community, as well as provide an indication on the clinical usefulness of the methodology. The best-known website with interactive clinical prediction tools is Adjuvant! Online.<sup>184</sup> This website provides decision support for adjuvant therapy (for example, chemotherapy and hormone therapy) after surgery for patients with early-stage cancer. Many researchers have evaluated the models available on this prediction website, thereby refining them with additional predictors and updated external validations.<sup>185,186</sup> A prediction website that focuses on decision support for radiotherapy was recently established.<sup>187</sup> The aim of this website is to let users work with and validate the interactive models developed for patients with cancer treated with radiotherapy, which contributes to CDSS development in general by demonstrating the potential of these predictions and raising the awareness of their existence and limitations.

## Future prospects

The major focus of this Review, thus far, has been model development, validation and presentation (including the features from different domains that might be considered as predictive and prognostic). Although an accurate outcome prediction model forms the basis of a CDSS, additional considerations must be made before a new CDSS can be used in daily radiation oncology practice.

First, any decision a patient or physician makes is based on a balance between its benefits (survival, local control and quality of life) and harms (toxic effects, complications, quality of life and financial cost). For example, an increased radiation dose usually results in both a higher probability of tumor control, but a concomitant higher probability of normal-tissue complications. Identifying the right balance between harm and benefit is a deeply personal choice that can vary substantially among patients. Thus, a CDSS should simultaneously predict local control, survival, treatment toxicity, quality of life and cost. The system should represent these

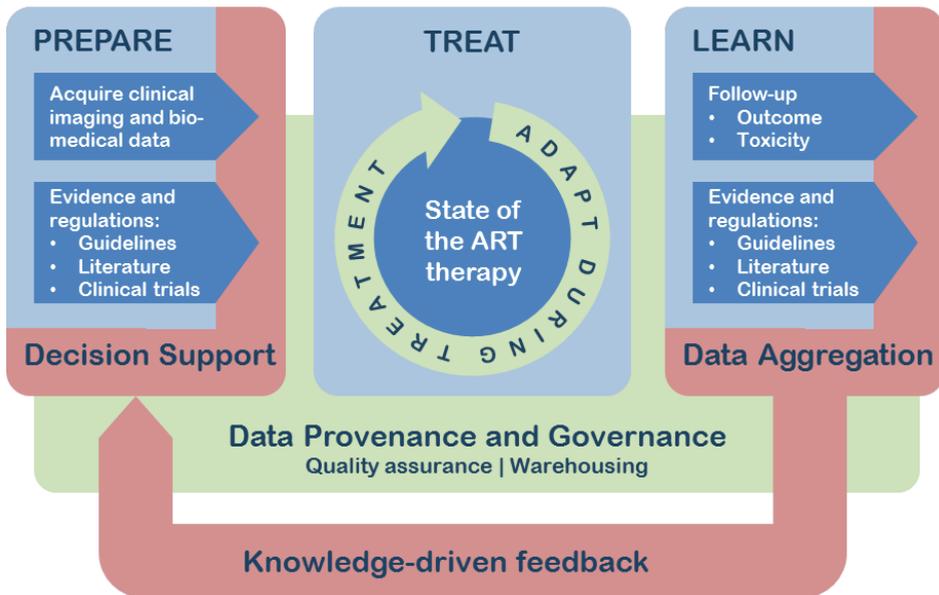
predictions and the balance between them in a way that is not only clear to the physician, but also to the patient, to achieve shared decision making.

Additionally, any prediction using a CDSS should be accompanied by a confidence interval. Accurately evaluating the confidence interval is an active and challenging area of research because uncertainties in the input features, missing features, size and quality of the training set and the intrinsic uncertainty of cancer must be incorporated to specify the uncertainty in the prediction for an individual patient. Without knowing if two possible decisions have a statistically significant and clinically meaningful difference in outcome, clinical decision support is difficult. Always sharing the data on which the model was based is a crucial prerequisite for this effort.

Current prediction models for decision support can only assist in distinguishing very high-level decisions—such as palliative versus curative treatment, sequential versus concurrent chemoradiation, surgery versus a wait-and-see approach. The radiation oncology community, however, is probably more interested in decisions such as intensity-modulated radiotherapy versus 3D-conformal radiotherapy or accelerated versus non-accelerated treatment. The current prediction models are simply not trained on datasets with these detailed subgroups and are not, therefore, accurate enough to support these decisions. Whether learning from increasingly diverse patient groups and adding other features will sufficiently improve the current models is unclear. As a result, tightly controlled studies using evidence-based medicine approaches are still crucial to guide clinical practice.

Finally, CDSSs should be seen as medical devices that require stringent acceptance, commissioning and quality assurance by the local institute. The key part of the commissioning and subsequent quality assurance is to validate the accuracy of the prediction model in the local patient population. Indeed, local patient data should be collected and the predicted outcomes compared with actual outcomes to convince local physicians that the support system works in their local setting. This ‘local validation’ should be done at the commissioning stage, but should be repeated to ensure the decision support remains valid, despite changes in local practice. Validation studies need to indicate what will be the required commission frequency.

This required quality assurance also enables the improvement of the system as more patient data becomes available. Using routine patient data to extract knowledge and apply that knowledge immediately is called ‘rapid learning’.<sup>3,188</sup> Rapid learning via continuously updated CDSSs offers a way to quickly learn from retrospective data and include new datasets (such as randomized controlled trial results) to adapt treatment protocols and deliver personalized decision support.

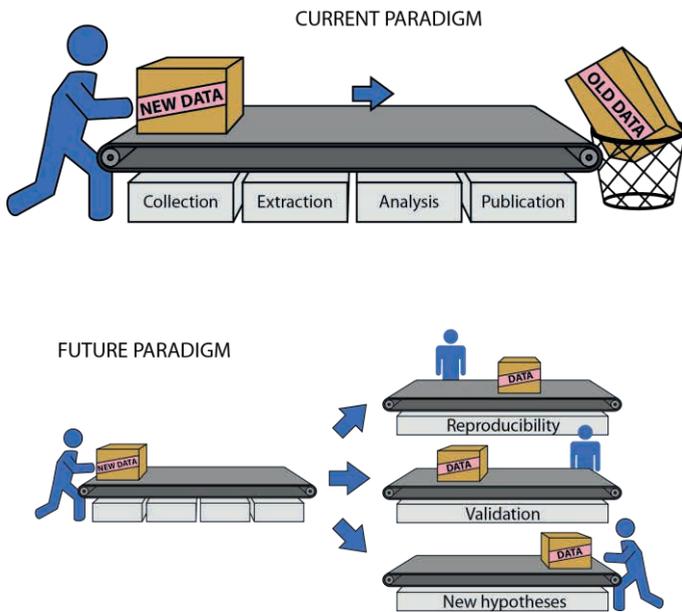


**Figure 2.6** Knowledge-driven health-care principles using a clinical decision support system in conjunction with standard evidence and regulations to choose the optimal treatment. In learning from follow-up data, knowledge is fed back to improve the clinical decision support system and adapt regulations.

As a data-driven discipline with well-established standards, such as DICOM-RT (digital imaging and communications in medicine in radiotherapy), radiotherapy offers an excellent starting point for adopting these rapid-learning principles (Figure 2.6). Aside from the importance of local data capture, which is still often lacking for (patient-reported) outcome and toxicity in particular, the quantity and heterogeneity of data that is necessary for rapid learning requires the pooling of data in a multi-institutional, international fashion.<sup>189,190</sup> One method of pooling data is to replicate routine clinical data sources in a distributed de-identified data warehouse, such as what is done in an international Computer-Aided Theragnostics network.<sup>191</sup> Examples of initiatives that create large centralized data and tissue infrastructures for routine radiation oncology patients are GENEPI,<sup>192</sup> the Radiogenomics Consortium,<sup>193</sup> ALLEGRO<sup>58</sup> and ULICE<sup>194</sup>. These initiatives also facilitate studies for external validation, reproducibility and hypothesis generation (Figure 2.7).<sup>189</sup>

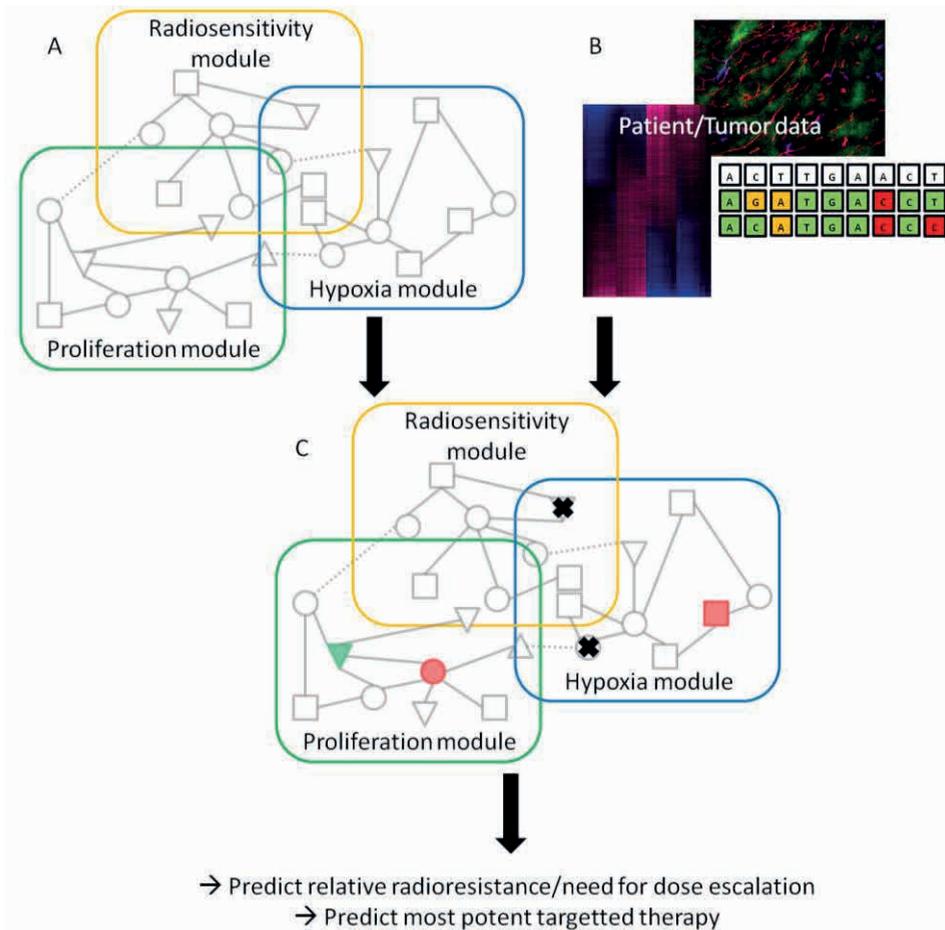
As datasets become larger (both in number of patients and in number of features per patient) high-throughput methods, both molecular<sup>195-200</sup> and imaging-based,<sup>79</sup> can produce large numbers of features that correlate with outcome.<sup>68,70,201-203</sup> A limited

application of these techniques has already transformed our understanding of radiotherapy response. For example, GWAS have associated SNPs with radiation toxicity.<sup>204,205</sup> Similarly, mRNA-abundance microarrays have been used to predict tumor response and normal-tissue toxicity in both patient and *in vitro* studies,<sup>206-210</sup> as well as to create markers that reflect biological phenotypes that are important for radiation response, such as hypoxia<sup>211,212</sup> and proliferation.<sup>213</sup> Both the data analysis and validation are important but challenging aspects of model development.<sup>195,214</sup> For example, the studies described above suffer from the substantial multiple-testing problem (that is, a large number of measured features compared with the sample number), which renders their results preliminary.<sup>206-213</sup> Human input and large, robust validation studies are, therefore, needed before features from high-throughput techniques can be included in CDSSs.<sup>215-217</sup>



**Figure 2.7** Illustration of the data handling paradigms. In the current paradigm the datasets are lost to the scientific community after collection, analysis and publication. In the future paradigm all the data will be shared in a standardized way and reused for e.g. reproducibility studies, validation studies, and studies testing new hypotheses. Pooling of multiple datasets may also improve accuracy of future predictions. This illustration is inspired on Deasy et al.<sup>189</sup>

Although studies on a single feature can be informative, only its combination into multimodal, multivariate models can be expected to provide a more holistic view of the response to radiation.



**Figure 2.8** A simplified schematic representation of systems biology applied to radiotherapy. (A) On the basis of in-vitro, in-vivo and patient data, modules representing the three biological categories (gene expression, immunohistochemical data and mutation data) important for radiotherapy response can be created. (B) For an individual patient, appropriate molecular data will be accumulated. (C) Combining the individual patient data with the modules will provide knowledge on specific module alterations (such as a deletion [X], upregulation [red] or downregulation [blue]), which can be translated to information on relative radioresistance and the molecular ‘weak’ spots of the tumor. This information will subsequently indicate whether dose escalation is necessary and which targeted drug is most effective for the patient. Part b used with permission from the National Academy of Sciences © Dubois, L. J. Proc. Natl Acad. Sci. USA 108, 14620–14625 (2011).

By combining events at different levels using systems-biology-like approaches, creating tumor-specific and patient-specific models of the effects and implications of radiation therapy should become possible (Figure 2.8).

Indeed, future studies will not only need to identify the individual components related to radiation response, but will also need to establish the interactions and relations amongst them.<sup>218</sup> Although this approach has not yet been applied to model radiotherapy responses, at least one study has demonstrated that combining multiple high-throughput data types can be used to map molecular cancer characteristics.<sup>219</sup> Combining models at different levels (societal, patient, whole tumor or organ, local tumor or organ, and cellular) is expected to lead to an increasingly holistic and accurate CDSS for the individual patient. Evidence that longitudinal data have added value to predicting outcome in, for example, repeated PET-imaging<sup>220</sup> and tumor perfusion<sup>221</sup> studies is growing, implying that this data need to be taken into account as candidates for future CDSSs.

Despite the challenges that remain, the vision of predictive models leading to CDSSs that are continuously updated via rapid learning on large datasets is clear, and numerous steps have already been taken. These include universal data-quality assurance programs and semantic interoperability issues.<sup>222</sup> However, we believe that this truly innovative journey will lead to necessary improvement of healthcare effectiveness and efficiency. Indeed, investments are being made in research and innovation for health-informatics systems, with an emphasis on interoperability and standards for secured data transfer, which shows that ‘eHealth’ will be among the largest healthcare innovations of the coming decade.<sup>222,223</sup>

## Conclusions

Accurate, externally validated prediction models are being rapidly developed, whereby multiple features related to the patient’s disease are combined into an integrated prediction. The key, however, is standardization—mainly in data acquisition across all areas, including molecular-based and imaging-based assays, patient preferences and possible treatments. Standardization requires harmonized clinical guidelines, regulated image acquisition and analysis parameters, validated biomarker assay criteria and data-sharing methods that use identical ontologies. Assessing the clinical usefulness of any CDSS is just as important as standardizing the development of externally validated accurate prediction models with high-quality data, preferably by standardizing the design of clinical trials. These crucial steps are the basis of validating a CDSS, which, in turn, will stimulate developments in rapid-learning healthcare and will enable the next major advances in shared decision making.

***Review criteria***

The PubMed and MEDLINE databases were searched for articles published in English (not restricted by date of publication) using a range of key phrases, including but not limited to: “PET imaging”, “heterogeneity in imaging”, “tumor response in radiotherapy”.

## Reference List

1. Vogelzang NJ, Benowitz SI, Adams S, et al. Clinical cancer advances 2011: Annual Report on Progress Against Cancer from the American Society of Clinical Oncology. *J Clin Oncol* 2012; 30:88-109
2. Fraass BA, Moran JM. Quality, technology and outcomes: evolution and evaluation of new treatments and/or new technology. *Semin Radiat Oncol* 2012; 22:3-10
3. Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. *J Clin Oncol* 2010; 28:4268-74
4. Maitland ML, Schilsky RL. Clinical trials in the era of personalized oncology. *CA Cancer J Clin* 2011; 61:365-81
5. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012; 366:883-92
6. Bachtiry B, Boutros PC, Pintilie M, et al. Gene expression profiling in cervical cancer: an exploration of intratumor heterogeneity. *Clin Cancer Res* 2006; 12:5632-40
7. Boyd CA, Benarroch-Gampel J, Sheffield KM, et al. 415 patients with adenocarcinoma of the pancreas: a population-based analysis of prognosis and survival. *J Surg Res* 2012; 174:12-9
8. Milosevic MF, Fyles AW, Wong R, et al. Interstitial fluid pressure in cervical carcinoma: within tumor heterogeneity, and relation to oxygen tension. *Cancer* 1998; 82:2418-26
9. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 2012;
10. Suit H, Skates S, Taghian A, et al. Clinical implications of heterogeneity of tumor response to radiation therapy. *Radiother Oncol* 1992; 25:251-60
11. Aerts HJ, Bussink J, Oyen WJ, et al. Identification of residual metabolic-active areas within NSCLC tumours using a pre-radiotherapy FDG-PET-CT scan: a prospective validation. *Lung Cancer* 2012; 75:73-6
12. Aerts HJ, van Baardwijk AA, Petit SF, et al. Identification of residual metabolic-active areas within individual NSCLC tumours using a pre-radiotherapy (18)Fluorodeoxyglucose-PET-CT scan. *Radiother Oncol* 2009; 91:386-92
13. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than harm? *J Clin Oncol* 2011; 29:2951-2
14. Bright TJ, Wong A, Dhurjati R, et al. Effect of Clinical Decision-Support Systems: A Systematic Review. *Ann Intern Med* 2012;
15. Clark GM. Prognostic factors versus predictive factors: Examples from a clinical trial of erlotinib. *Mol Oncol* 2008; 1:406-12
16. Dancey JE, Dobbin KK, Groshen S, et al. Guidelines for the development and incorporation of biomarker studies in early clinical trials of novel agents. *Clin Cancer Res* 2010; 16:1745-55
17. Peek N, Arts DG, Bosman RJ, et al. External validation of prognostic models for critically ill patients required substantial sample sizes. *J Clin Epidemiol* 2007; 60:491-501
18. Vergouwe Y, Steyerberg EW, Eijkemans MJ, et al. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005; 58:475-83
19. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; 21:128-38
20. Aittokallio T. Dealing with missing values in large-scale studies: microarray data imputation and beyond. *Brief Bioinform* 2010; 11:253-64

21. Ludbrook J. Outlying observations and missing values: how should they be handled? *Clin Exp Pharmacol Physiol* 2008; 35:670-8
22. Jayalakshmi T, Santhakumaran A. Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering* 2011; 3:89-93
23. Huan L, Motoda H: *Feature Selection for Knowledge Discovery and Data Mining*, 1998, Springer
24. Harrell FEJ: *Regression modeling strategies*, 2001, Springer, New York
25. Bishop CM: *Pattern Recognition and Machine Learning*, 1st ed, 2007, Springer
26. Lee SM, Abbott PA. Bayesian networks for knowledge discovery in large datasets: basics for nurse researchers. *J Biomed Inform* 2003; 36:389-99
27. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2007; 2:59-77
28. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007; 26:2389-430
29. Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; 98:691-8
30. Dehing-Oberije C, De Ruyscher D, Petit S, et al. Development, external validation and clinical usefulness of a practical prediction model for radiation-induced dysphagia in lung cancer patients. *Radiother Oncol* 2010; 97:455-61
31. Specht MC, Kattan MW, Gonen M, et al. Predicting nonsentinel node status after positive sentinel lymph biopsy for breast cancer: clinicians versus nomogram. *Ann Surg Oncol* 2005; 12:654-9
32. Dehing-Oberije C, De Ruyscher D, van der Weide H, et al. Tumor volume combined with number of positive lymph node stations is a more important prognostic factor than TNM stage for survival of non-small-cell lung cancer patients treated with (chemo)radiotherapy. *Int J Radiat Oncol Biol Phys* 2008; 70:1039-44
33. Vickers AJ, Kramer BS, Baker SG. Selecting patients for randomized trials: a systematic approach based on risk group. *Trials* 2006; 7:30
34. Baggerly KA, Coombes KR. What Information Should Be Required to Support Clinical "Omics" Publications? *Clin Chem* 2011;
35. Klopp AH, Eifel PJ. Biological predictors of cervical cancer response to radiation therapy. *Semin Radiat Oncol* 2012; 22:143-50
36. Kristiansen G. Diagnostic and prognostic molecular biomarkers for prostate cancer. *Histopathology* 2012; 60:125-41
37. Dehing-Oberije C, Yu S, De Ruyscher D, et al. Development and external validation of prognostic model for 2-year survival of non-small-cell lung cancer patients treated with chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2009; 74:355-62
38. Ang CS, Phung J, Nice EC. The discovery and validation of colorectal cancer biomarkers. *Biomed Chromatogr* 2011; 25:82-99
39. Schmidt ME, Steindorf K. Statistical methods for the validation of questionnaires--discrepancy between theory and practice. *Methods Inf Med* 2006; 45:409-13
40. Garrido-Laguna I, Janku F, Vaklavas C, et al. Validation of the Royal Marsden Hospital prognostic score in patients treated in the Phase I Clinical Trials Program at the MD Anderson Cancer Center. *Cancer* 2012; 118:1422-8
41. Shrier I, Boivin JF, Steele RJ, et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; 166:1203-9

42. Tzoulaki I, Siontis KC, Ioannidis JP. Prognostic effect size of cardiovascular biomarkers in datasets from observational studies versus randomised trials: meta-epidemiology study. *BMJ* 2011; 343:d6829
43. Trotti A, Colevas AD, Setser A, et al. Patient-reported outcomes and the evolution of adverse event reporting in oncology. *J Clin Oncol* 2007; 25:5121-7
44. Trotti A, Colevas AD, Setser A, et al. CTCAE v3.0: development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol* 2003; 13:176-81
45. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med* 2011; 8:e1001026
46. Steg PG, Lopez-Sendon J, Lopez de Sa E, et al. External validity of clinical trials in acute myocardial infarction. *Arch Intern Med* 2007; 167:68-73
47. Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association studies (STREGA): an extension of the STROBE Statement. *Ann Intern Med* 2009; 150:206-15
48. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; 370:1453-7
49. Dawson LA, Sharpe MB. Image-guided radiotherapy: rationale, benefits, and limitations. *Lancet Oncol* 2006; 7:848-58
50. Bentzen SM, Constine LS, Deasy JO, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): an introduction to the scientific issues. *Int J Radiat Oncol Biol Phys* 2010; 76:S3-9
51. Verellen D, De Ridder M, Linthout N, et al. Innovations in image-guided radiotherapy. *Nat Rev Cancer* 2007; 7:949-60
52. Holthusen H. Erfahrungen über die Verträglichkeitsgrenze für Röntgenstrahlen und deren Nutzenanwendung zur Verhütung von Schäden. *Strahlentherapie*. 1936; 57:254-269
53. Valentini V, van Stiphout RG, Lammering G, 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. *J Clin Oncol* 2011; 29:3163-72
54. Belderbos J, Uitterhoeve L, van Zandwijk N, et al. Randomised trial of sequential versus concurrent chemo-radiotherapy in patients with inoperable non-small cell lung cancer (EORTC 08972-22973). *Eur J Cancer* 2007; 43:114-21
55. Lambin P, Petit SF, Aerts HJ, 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. *Radiother Oncol* 2010; 96:145-52
56. Graham MV, Purdy JA, Emami B, et al. Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC). *Int J Radiat Oncol Biol Phys* 1999; 45:323-9
57. Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation. *Int J Radiat Oncol Biol Phys* 1991; 21:109-22
58. Ottolenghi A, Smyth V, Trott KR. The risks to healthy tissues from the use of existing and emerging techniques for radiation therapy. *Radiat Prot Dosimetry* 2011; 143:533-5
59. Beetz I, Schilstra C, van der Schaaf A, et al. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: The role of dosimetric and clinical factors. *Radiother Oncol* 2012;
60. van der Schaaf A, Xu CJ, van Luijk P, et al. Multivariate modeling of complications with data driven variable selection: Guarding against overfitting and effects of data set size. *Radiother Oncol* 2012;
61. Xu CJ, van der Schaaf A, Van't Veld AA, et al. Statistical validation of normal tissue complication probability models. *Int J Radiat Oncol Biol Phys* 2012; 84:e123-9

62. Nijsten SM, Mijnheer BJ, Dekker AL, et al. Routine individualised patient dosimetry using electronic portal imaging devices. *Radiother Oncol* 2007; 83:65-75
63. van Elmpt W, Petit S, De Ruyscher D, et al. 3D dose delivery verification using repeated cone-beam imaging and EPID dosimetry for stereotactic body radiotherapy of non-small cell lung cancer. *Radiother Oncol* 2010; 94:188-94
64. van Elmpt W, Nijsten S, Petit S, et al. 3D in vivo dosimetry using megavoltage cone-beam CT and EPID dosimetry. *Int J Radiat Oncol Biol Phys* 2009; 73:1580-7
65. Rodrigues G, Lock M, D'Souza D, et al. Prediction of radiation pneumonitis by dose - volume histogram parameters in lung cancer--a systematic review. *Radiother Oncol* 2004; 71:127-38
66. Werner-Wasik M, Yorke E, Deasy J, et al. Radiation dose-volume effects in the esophagus. *Int J Radiat Oncol Biol Phys* 2010; 76:S86-93
67. Saunders M, Rojas AM, Dische S. CHART revisited: a conservative approach for advanced head and neck cancer. *Clin Oncol (R Coll Radiol)* 2008; 20:127-33
68. Turner N, Lambros MB, Horlings HM, et al. Integrative molecular profiling of triple negative breast cancers identifies amplicon drivers and potential therapeutic targets. *Oncogene* 2010; 29:2013-23
69. Begg AC, Haustermans K, Hart AA, et al. The value of pretreatment cell kinetic parameters as predictors for radiotherapy outcome in head and neck cancer: a multicenter analysis. *Radiother Oncol* 1999; 50:13-23
70. Taguchi F, Solomon B, Gregorc V, et al. Mass spectrometry to classify non-small-cell lung cancer patients for clinical outcome after treatment with epidermal growth factor receptor tyrosine kinase inhibitors: a multicohort cross-institutional study. *J Natl Cancer Inst* 2007; 99:838-46
71. Hessel F, Petersen C, Zips D, et al. Impact of increased cell loss on the repopulation rate during fractionated irradiation in human FaDu squamous cell carcinoma growing in nude mice. *Int J Radiat Biol* 2003; 79:479-86
72. Baumann M, Krause M, Hill R. Exploring the role of cancer stem cells in radioresistance. *Nat Rev Cancer* 2008; 8:545-54
73. Ben-Josef E, Moughan J, Ajani JA, et al. Impact of overall treatment time on survival and local control in patients with anal cancer: a pooled data analysis of Radiation Therapy Oncology Group trials 87-04 and 98-11. *J Clin Oncol* 2010; 28:5061-6
74. Thames HD, Kuban D, Levy LB, et al. The role of overall treatment time in the outcome of radiotherapy of prostate cancer: an analysis of biochemical failure in 4839 men treated between 1987 and 1995. *Radiother Oncol* 2010; 96:6-12
75. Leonard F. Imaging and cancer: A review. *Molecular Oncology* 2008; 2:115-152
76. Torigian DA, Huang SS, Houseni M, et al. Functional imaging of cancer with emphasis on molecular techniques. *CA Cancer J Clin* 2007; 57:206-24
77. Eadie LH, Taylor P, Gibson AP. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *European Journal of Radiology* 2012; 81:e70-e76
78. Gillies RJ, Anderson AR, Gatenby RA, et al. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clinical Radiology* 2010; 65:517-521
79. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer* 2012; 48:441-446
80. Velazquez ER, Aerts HJ, Oberije C, et al. Prediction of residual metabolic activity after treatment in NSCLC patients. *Acta Oncol* 2010; 49:1033-9
81. Cangir AK, Kutlay H, Akal M, et al. Prognostic value of tumor size in non-small cell lung cancer larger than five centimeters in diameter. *Lung Cancer* 2004; 46:325-331

82. Lam JS, Klatte T, Patard J-J, et al. Prognostic Relevance of Tumour Size in T3a Renal Cell Carcinoma: A Multicentre Experience. *European Urology* 2007; 52:155-162
83. Pitson G, Fyles A, Milosevic M, et al. Tumor size and oxygenation are independent predictors of nodal diseases in patients with cervix cancer. *International Journal of Radiation Oncology\*Biography\*Physics* 2001; 51:699-703
84. Thomas F, Arriagada R, Mouriessse H, et al. Radical radiotherapy alone in non-operable breast cancer: The major impact of tumor size and histological grade on prognosis. *Radiotherapy and Oncology* 1988; 13:267-276
85. Steenbakkers RJ, Duppen JC, Fitton I, et al. Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: a 'Big Brother' evaluation. *Radiother Oncol* 2005; 77:182-90
86. Greco C, Rosenzweig K, Cascini GL, et al. Current status of PET/CT for tumour volume definition in radiotherapy treatment planning for non-small cell lung cancer (NSCLC). *Lung Cancer* 2007; 57:125-34
87. Caldwell CB, Mah K, Ung YC, et al. Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: the impact of 18FDG-hybrid PET fusion. *Int J Radiat Oncol Biol Phys* 2001; 51:923-31
88. Bowden P, Fisher R, Mac Manus M, et al. Measurement of lung tumor volumes using three-dimensional computer planning software. *Int J Radiat Oncol Biol Phys* 2002; 53:566-73
89. Nishino M, Guo M, Jackman DM, et al. CT tumor volume measurement in advanced non-small-cell lung cancer: Performance characteristics of an emerging clinical tool. *Acad Radiol* 2011; 18:54-62
90. Marcus CD, Ladam-Marcus V, Cucu C, et al. Imaging techniques to evaluate the response to treatment in oncology: current standards and perspectives. *Crit Rev Oncol Hematol* 2009; 72:217-38
91. Schwartz LH, Mazumdar M, Brown W, et al. Variability in response assessment in solid tumors: effect of number of lesions chosen for measurement. *Clin Cancer Res* 2003; 9:4318-23
92. Erasmus JJ, Gladish GW, Broemeling L, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol* 2003; 21:2574-82
93. Therasse P. Measuring the clinical response. What does it mean? *Eur J Cancer* 2002; 38:1817-23
94. Nehmeh SA, Erdi YE. Respiratory motion in positron emission tomography/computed tomography: a review. *Semin Nucl Med* 2008; 38:167-76
95. Sonke JJ, Belderbos J. Adaptive radiotherapy for lung cancer. *Semin Radiat Oncol* 2010; 20:94-106
96. van Baardwijk A, Bosmans G, Boersma L, et al. PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *Int J Radiat Oncol Biol Phys* 2007; 68:771-8
97. Wu K, Ung YC, Hornby J, et al. PET CT thresholds for radiotherapy target definition in non-small-cell lung cancer: how close are we to the pathologic findings? *Int J Radiat Oncol Biol Phys* 2010; 77:699-706
98. Wanet M, Lee JA, Weynand B, et al. Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens. *Radiother Oncol* 2011; 98:117-25
99. Strassmann G, Abdellaoui S, Richter D, et al. Atlas-based semiautomatic target volume definition (CTV) for head-and-neck tumors. *Int J Radiat Oncol Biol Phys* 2010; 78:1270-6
100. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med* 2005; 46:1342-8

101. Daisne JF, Duprez T, Weynand B, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology* 2004; 233:93-100
102. van Loon J, van Baardwijk A, Boersma L, et al. Therapeutic implications of molecular imaging with PET in the combined modality treatment of lung cancer. *Cancer Treat Rev* 2011; 37:331-43
103. Wood KA, Hoskin PJ, Saunders MI. Positron Emission Tomography in Oncology: A Review. *Clinical Oncology* 2007; 19:237-255
104. O'Connor JPB, Jackson A, Asselin M-C, et al. Quantitative imaging biomarkers in the clinical development of targeted therapeutics: current and future perspectives. *The Lancet Oncology* 2008; 9:766-776
105. van Baardwijk A, Bosmans G, Dekker A, et al. Time trends in the maximal uptake of FDG on PET scan during thoracic radiotherapy. A prospective study in locally advanced non-small cell lung cancer (NSCLC) patients. *Radiotherapy and Oncology* 2007; 82:145-152
106. Rodney J H. PET for Therapeutic Response Monitoring in Oncology. *PET Clinics* 2008; 3:89-99
107. Chung HH, Kim JW, Han KH, et al. Prognostic value of metabolic tumor volume measured by FDG-PET/CT in patients with cervical cancer. *Gynecologic Oncology* 2011; 120:270-274
108. Borst GR, Belderbos JSA, Boellaard R, et al. Standardised FDG uptake: A prognostic factor for inoperable non-small cell lung cancer. *European Journal of Cancer* 2005; 41:1533-1541
109. Mac Manus MP, Hicks RJ, Matthews JP, et al. Metabolic (FDG-PET) response after radical radiotherapy/chemoradiotherapy for non-small cell lung cancer correlates with patterns of failure. *Lung Cancer* 2005; 49:95-108
110. Hoekstra CJ, Stroobants SG, Smit EF, et al. Prognostic relevance of response evaluation using [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. *J Clin Oncol* 2005; 23:8362-70
111. Soto DE, Kessler ML, Piert M, et al. Correlation between pretreatment FDG-PET biological target volume and anatomical location of failure after radiation therapy for head and neck cancers. *Radiother Oncol* 2008; 89:13-8
112. Lambrecht M, Deroose C, Roels S, et al. The use of FDG-PET/CT and diffusion-weighted magnetic resonance imaging for response prediction before, during and after preoperative chemoradiotherapy for rectal cancer. *Acta Oncol* 2010; 49:956-63
113. Janssen MHM, Ållers MC, van Stiphout RGPM, et al. Evaluation of early metabolic responses in rectal cancer during combined radiochemotherapy or radiotherapy alone: Sequential FDG-PET-CT findings. *Radiotherapy and Oncology* 2010; 94:151-155
114. Ceulemans Gt, Voordeckers M, Farrag A, et al. Can 18-FDG-PET During Radiotherapy Replace Post-Therapy Scanning for Detection/Demonstration of Tumor Response in Head-and-Neck Cancer? *International Journal of Radiation Oncology\*Biophysics* 2011; 81:938-942
115. van Loon J, Offermann C, Ollers M, et al. Early CT and FDG-metabolic tumour volume changes show a significant correlation with survival in stage I-III small cell lung cancer: a hypothesis generating study. *Radiother Oncol* 2011; 99:172-5
116. Bussink J, Kaanders JH, van der Graaf WT, et al. PET-CT for radiotherapy treatment planning and response monitoring in solid tumors. *Nat Rev Clin Oncol* 2011; 8:233-42
117. Boellaard R. Need for standardization of 18F-FDG PET/CT for treatment response assessments. *J Nucl Med* 2011; 52 Suppl 2:93S-100S
118. Boellaard R, Oyen WJ, Hoekstra CJ, et al. The Netherlands protocol for standardisation and quantification of FDG whole body PET studies in multi-centre trials. *Eur J Nucl Med Mol Imaging* 2008; 35:2320-33

119. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging* 2010; 37:181-200
120. Bayouth JE, Casavant TL, Graham MM, et al. Image-Based Biomarkers in Clinical Practice. *Seminars in Radiation Oncology* 2011; 21:157-166
121. Harry VN, Semple SI, Parkin DE, et al. Use of new imaging techniques to predict tumour response to therapy. *The Lancet Oncology* 2010; 11:92-102
122. Heijmen L, Verstappen MC, Ter Voert EE, et al. Tumour response prediction by diffusion-weighted MR imaging: Ready for clinical use? *Crit Rev Oncol Hematol* 2012;
123. Lambrecht M, Vandecaveye V, Van Limbergen EJ, et al. The prognostic value of pretherapeutic diffusion-weighted MRI in oropharyngeal carcinoma treated with (chemo-)radiotherapy. *Cancer Imaging* 2011; 11 Spec No A:S112-3
124. Vandecaveye V, Dirix P, De Keyzer F, et al. Diffusion-weighted magnetic resonance imaging early after chemoradiotherapy to monitor treatment response in head-and-neck squamous cell carcinoma. *Int J Radiat Oncol Biol Phys* 2012; 82:1098-107
125. Kim SY, Lee SS, Byun JH, et al. Malignant hepatic tumors: short-term reproducibility of apparent diffusion coefficients with breath-hold and respiratory-triggered diffusion-weighted MR imaging. *Radiology* 2010; 255:815-23
126. Sinkus R, Van Beers BE, Vilgrain V, et al. Apparent diffusion coefficient from magnetic resonance imaging as a biomarker in oncology drug development. *Eur J Cancer* 2012; 48:425-31
127. Kierkels RGJ, Backes WH, Janssen MHM, et al. Comparison Between Perfusion Computed Tomography and Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Rectal Cancer. *International Journal of Radiation Oncology\*Biography\*Physics* 2010; 77:400-408
128. Shukla-Dave A, Lee NY, Jansen JFA, et al. Dynamic Contrast-Enhanced Magnetic Resonance Imaging as a Predictor of Outcome in Head and Neck Squamous Cell Carcinoma Patients with Nodal Metastases. *International Journal of Radiation Oncology\*Biography\*Physics* 2012;
129. Yaromina A, Quennet V, Zips D, et al. Co-localisation of hypoxia and perfusion markers with parameters of glucose metabolism in human squamous cell carcinoma (hSCC) xenografts. *Int J Radiat Biol* 2009; 85:972-80
130. Morchel P, Melkus G, Yaromina A, et al. Correlating quantitative MR measurements of standardized tumor lines with histological parameters and tumor control dose. *Radiother Oncol* 2010; 96:123-30
131. Quennet V, Yaromina A, Zips D, et al. Tumor lactate content predicts for response to fractionated irradiation of human squamous cell carcinomas in nude mice. *Radiother Oncol* 2006; 81:130-5
132. Kim YI, Chung JW, Park JH, et al. Multiphase contrast-enhanced CT imaging in hepatocellular carcinoma correlation with immunohistochemical angiogenic activities. *Acad Radiol* 2007; 14:1084-91
133. Miles KA. Perfusion CT for the assessment of tumour vascularity: which protocol? *Br J Radiol* 2003; 76 Spec No 1:S36-42
134. Miles KA. Molecular imaging with dynamic contrast-enhanced computed tomography. *Clinical Radiology* 2010; 65:549-556
135. Petralia G, Bonello L, Viotti S, et al. CT perfusion in oncology: how to do it. *Cancer Imaging* 2010; 10:8-19
136. Asselin MC, O'Connor JP, Boellaard R, et al. Quantifying heterogeneity in human tumours using MRI and PET. *Eur J Cancer* 2012; 48:447-55
137. Eary JF, O'Sullivan F, O'Sullivan J, et al. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med* 2008; 49:1973-9

138. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med* 2011; 52:369-78
139. Diehn M, Nardini C, Wang DS, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci U S A* 2008; 105:5213-8
140. Kuo MD, Gollub J, Sirlin CB, et al. Radiogenomic analysis to identify imaging phenotypes associated with drug response gene expression programs in hepatocellular carcinoma. *J Vasc Interv Radiol* 2007; 18:821-31
141. Segal E, Sirlin CB, Ooi C, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nat Biotechnol* 2007; 25:675-80
142. Rutman AM, Kuo MD. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. *Eur J Radiol* 2009; 70:232-41
143. Lindegaard JC, Overgaard J, Bentzen SM, et al. Is there a radiobiologic basis for improving the treatment of advanced stage cervical cancer? *J Natl Cancer Inst Monogr* 1996:105-12
144. Slonina D, Gasinska A. Intrinsic radiosensitivity of healthy donors and cancer patients as determined by the lymphocyte micronucleus assay. *Int J Radiat Biol* 1997; 72:693-701
145. Fertil B, Malaise EP. Intrinsic radiosensitivity of human cell lines is correlated with radioresponsiveness of human tumors: analysis of 101 published survival curves. *Int J Radiat Oncol Biol Phys* 1985; 11:1699-707
146. West CM, Davidson SE, Roberts SA, et al. The independence of intrinsic radiosensitivity as a prognostic factor for patient response to radiotherapy of carcinoma of the cervix. *Br J Cancer* 1997; 76:1184-90
147. Bjork-Eriksson T, West C, Karlsson E, et al. Tumor radiosensitivity (SF2) is a prognostic factor for local control in head and neck cancers. *Int J Radiat Oncol Biol Phys* 2000; 46:13-9
148. Bartelink H, Begg A, Martin JC, et al. Towards prediction and modulation of treatment response. *Radiother Oncol* 1999; 50:1-11
149. Begg AC. Predicting recurrence after radiotherapy in head and neck cancer. *Semin Radiat Oncol* 2012; 22:108-18
150. Menegakis A, Yaromina A, Eicheler W, et al. Prediction of clonogenic cell survival curves based on the number of residual DNA double strand breaks measured by gammaH2AX staining. *Int J Radiat Biol* 2009; 85:1032-41
151. Olive PL, Banath JP. Phosphorylation of histone H2AX as a measure of radiosensitivity. *Int J Radiat Oncol Biol Phys* 2004; 58:331-5
152. Hockel M, Schlenger K, Aral B, et al. Association between tumor hypoxia and malignant progression in advanced cancer of the uterine cervix. *Cancer Res* 1996; 56:4509-15
153. Vaupel P, Mayer A. Hypoxia in cancer: significance and impact on clinical outcome. *Cancer Metastasis Rev* 2007; 26:225-39
154. Chouaib S, Messai Y, Couve S, et al. Hypoxia promotes tumor growth in linking angiogenesis to immune escape. *Front Immunol* 2012; 3:21
155. Kaanders JH, Wiffjffels KI, Marres HA, et al. Pimonidazole binding and tumor vascularity predict for treatment outcome in head and neck cancer. *Cancer Res* 2002; 62:7066-74
156. Nordmark M, Loncaster J, Aquino-Parsons C, et al. The prognostic value of pimonidazole and tumour pO<sub>2</sub> in human cervix carcinomas after radiation therapy: a prospective international multi-center study. *Radiother Oncol* 2006; 80:123-31
157. Rouschop KM, van den Beucken T, Dubois L, et al. The unfolded protein response protects human tumor cells during hypoxia through regulation of the autophagy genes MAP1LC3B and ATG5. *J Clin Invest* 2010; 120:127-41

158. Krause BJ, Beck R, Souvatzoglou M, et al. PET and PET/CT studies of tumor tissue oxygenation. *Q J Nucl Med Mol Imaging* 2006; 50:28-43
159. Dubois LJ, Lieuwes NG, Janssen MH, et al. Preclinical evaluation and validation of [18F]HX4, a promising hypoxia marker for PET imaging. *Proc Natl Acad Sci U S A* 2011; 108:14620-5
160. van Loon J, De Ruyscher D, Wanders R, et al. Selective nodal irradiation on basis of (18)FDG-PET scans in limited-disease small-cell lung cancer: a prospective study. *Int J Radiat Oncol Biol Phys* 2010; 77:329-36
161. West CM, Cooper RA, Loncaster JA, et al. Tumor vascularity: a histological measure of angiogenesis and hypoxia. *Cancer Res* 2001; 61:2907-10
162. Maciejewski B, Withers HR, Taylor JM, et al. Dose fractionation and regeneration in radiotherapy for cancer of the oral cavity and oropharynx: tumor dose-response and repopulation. *Int J Radiat Oncol Biol Phys* 1989; 16:831-43
163. Turesson I, Nyman J, Holmberg E, et al. Prognostic factors for acute and late skin reactions in radiotherapy patients. *Int J Radiat Oncol Biol Phys* 1996; 36:1065-75
164. Johansen J, Bentzen SM, Overgaard J, et al. Evidence for a positive correlation between in vitro radiosensitivity of normal human skin fibroblasts and the occurrence of subcutaneous fibrosis after radiotherapy. *Int J Radiat Biol* 1994; 66:407-12
165. West CM, Davidson SE, Elyan SA, et al. Lymphocyte radiosensitivity is a significant prognostic factor for morbidity in carcinoma of the cervix. *Int J Radiat Oncol Biol Phys* 2001; 51:10-5
166. Peacock J, Ashton A, Bliss J, et al. Cellular radiosensitivity and complication risk after curative radiotherapy. *Radiother Oncol* 2000; 55:173-8
167. Russell NS, Grummels A, Hart AA, et al. Low predictive value of intrinsic fibroblast radiosensitivity for fibrosis development following radiotherapy for breast cancer. *Int J Radiat Biol* 1998; 73:661-70
168. Russell NS, Arlett CF, Bartelink H, et al. Use of fluorescence in situ hybridization to determine the relationship between chromosome aberrations and cell survival in eight human fibroblast strains. *Int J Radiat Biol* 1995; 68:185-96
169. Kiltie AE, Ryan AJ, Swindell R, et al. A correlation between residual radiation-induced DNA double-strand breaks in cultured fibroblasts and late radiotherapy reactions in breast cancer patients. *Radiother Oncol* 1999; 51:55-65
170. Dileto CL, Travis EL. Fibroblast radiosensitivity in vitro and lung fibrosis in vivo: comparison between a fibrosis-prone and fibrosis-resistant mouse strain. *Radiat Res* 1996; 146:61-7
171. Azria D, Belkacemi Y, Romieu G, et al. Concurrent or sequential adjuvant letrozole and radiotherapy after conservative surgery for early-stage breast cancer (CO-HO-RT): a phase 2 randomised trial. *Lancet Oncol* 2010; 11:258-65
172. Bentzen SM. Preventing or reducing late side effects of radiation therapy: radiobiology meets molecular pathology. *Nat Rev Cancer* 2006; 6:702-13
173. Rodemann HP, Bamberg M. Cellular basis of radiation-induced fibrosis. *Radiother Oncol* 1995; 35:83-90
174. Andreassen CN, Alsner J, Overgaard M, et al. Risk of radiation-induced subcutaneous fibrosis in relation to single nucleotide polymorphisms in TGFB1, SOD2, XRCC1, XRCC3, APEX and ATM—a study based on DNA from formalin fixed paraffin embedded tissue samples. *Int J Radiat Biol* 2006; 82:577-86
175. Chang-Claude J, Popanda O, Tan XL, et al. Association between polymorphisms in the DNA repair genes, XRCC1, APE1, and XPD and acute side effects of radiotherapy in breast cancer patients. *Clin Cancer Res* 2005; 11:4802-9
176. Barnett GC, Coles CE, Elliott RM, et al. Independent validation of genes and polymorphisms reported to be associated with radiation toxicity: a prospective analysis study. *Lancet Oncol* 2012; 13:65-77

177. Cammann H, Jung K, Meyer HA, et al. Avoiding pitfalls in applying prediction models, as illustrated by the example of prostate cancer diagnosis. *Clin Chem* 2011; 57:1490-8
178. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008; 26:1364-70
179. Dehing-Oberije C, Aerts H, Yu S, et al. Development and validation of a prognostic model using blood biomarker information for prediction of survival of non-small-cell lung cancer patients treated with combined chemotherapy and radiation or radiotherapy alone (NCT00181519, NCT00573040, and NCT00572325). *Int J Radiat Oncol Biol Phys* 2011; 81:360-8
180. Egelmeer AG, Velazquez ER, de Jong JM, et al. Development and validation of a nomogram for prediction of survival and local control in laryngeal carcinoma patients treated with radiotherapy alone: a cohort study based on 994 patients. *Radiother Oncol* 2011; 100:108-15
181. van Stiphout RG, Lammering G, Buijsen J, et al. Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential PET-CT imaging. *Radiother Oncol* 2011; 98:126-33
182. Marko NF, Xu Z, Gao T, et al. Predicting survival in women with breast cancer and brain metastasis: A nomogram outperforms current survival prediction models. *Cancer* 2011;
183. Rudloff U, Jacks LM, Goldberg JI, et al. Nomogram for predicting the risk of local recurrence after breast-conserving surgery for ductal carcinoma in situ. *J Clin Oncol* 2010; 28:3762-9
184. <http://www.adjuvantonline.com>,
185. Hajage D, de Rycke Y, Bollet M, et al. External validation of Adjuvant! Online breast cancer prognosis tool. Prioritising recommendations for improvement. *PLoS One* 2011; 6:e27446
186. Kuo YL, Chen DR, Chang TW. Accuracy validation of adjuvant! online in Taiwanese breast cancer patients - a 10-year analysis. *BMC Med Inform Decis Mak* 2012; 12:108
187. <http://www.predictcancer.org>,
188. Ginsburg GS, Staples J, Abernethy AP. Academic medical centers: ripe for rapid-learning personalized health care. *Sci Transl Med* 2011; 3:101cm27
189. Deasy JO, Bentzen SM, Jackson A, et al. Improving normal tissue complication probability models: the need to adopt a "data-pooling" culture. *Int J Radiat Oncol Biol Phys* 2010; 76:S151-4
190. Roelofs E, Persoon L, Qamhiyeh S, et al. Design of and technical challenges involved in a framework for multicentric radiotherapy treatment planning studies. *Radiother Oncol* 2010; 97:567-71
191. <http://www.eurocat.info>,
192. De Ruyscher D, Severin D, Barnes E, et al. First report on the patient database for the identification of the genetic pathways involved in patients over-reacting to radiotherapy: GENEPI-II. *Radiother Oncol* 2010; 97:36-9
193. West C, Rosenstein BS, Alsner J, et al. Establishment of a Radiogenomics Consortium. *Int J Radiat Oncol Biol Phys* 2010; 76:1295-6
194. Kessel KA, Bougatf N, Bohn C, et al. Connection of European particle therapy centers and generation of a common particle database system within the European ULICE-framework. *Radiat Oncol* 2012; 7:115
195. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 2010; 11:685-96
196. Wulfkuhle JD, Liotta LA, Petricoin EF. Proteomic applications for the early detection of cancer. *Nat Rev Cancer* 2003; 3:267-75
197. Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer. *Nat Genet* 2005; 37 Suppl:S11-7
198. Quackenbush J. Microarray analysis and tumor classification. *N Engl J Med* 2006; 354:2463-72

199. Lu Y, Muller M, Smith D, et al. Kinome siRNA-phosphoproteomic screen identifies networks regulating AKT signaling. *Oncogene* 2011; 30:4567-77
200. Gupta PB, Onder TT, Jiang G, et al. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 2009; 138:645-59
201. Shedden K, Taylor JM, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008; 14:822-7
202. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell* 2006; 9:189-98
203. Zuber J, Shi J, Wang E, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* 2011; 478:524-8
204. Best T, Li D, Skol AD, et al. Variants at 6q21 implicate PRDM1 in the etiology of therapy-induced second malignancies after Hodgkin's lymphoma. *Nat Med* 2011; 17:941-3
205. Kerns SL, Ostrer H, Stock R, et al. Genome-wide association study to identify single nucleotide polymorphisms (SNPs) associated with the development of erectile dysfunction in African-American men after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2010; 78:1292-300
206. de Jong MC, Pramana J, van der Wal JE, et al. CD44 expression predicts local recurrence after radiotherapy in larynx cancer. *Clin Cancer Res* 2010; 16:5329-38
207. Kitahara O, Katagiri T, Tsunoda T, et al. Classification of sensitivity or resistance of cervical cancers to ionizing radiation according to expression profiles of 62 genes selected by cDNA microarray analysis. *Neoplasia* 2002; 4:295-303
208. Torres-Roca JF, Eschrich S, Zhao H, et al. Prediction of radiation sensitivity using a gene expression classifier. *Cancer Res* 2005; 65:7169-76
209. Wong YF, Sahota DS, Cheung TH, et al. Gene expression pattern associated with radiotherapy sensitivity in cervical cancer. *Cancer J* 2006; 12:189-93
210. Yang S, Chen J, Guo Y, et al. Identification of prognostic biomarkers for response to radiotherapy by DNA microarray in nasopharyngeal carcinoma patients. *Int J Oncol* 2012; 40:1590-600
211. Starmans MH, Chu KC, Haider S, et al. The prognostic value of temporal in vitro and in vivo derived hypoxia gene-expression signatures in breast cancer. *Radiother Oncol* 2012; 102:436-43
212. Winter SC, Buffa FM, Silva P, et al. Relation of a hypoxia metagene derived from head and neck cancer to prognosis of multiple cancers. *Cancer Res* 2007; 67:3441-9
213. Starmans MH, Krishnapuram B, Steck H, et al. Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. *Br J Cancer* 2008; 99:1884-90
214. Koboldt DC, Ding L, Mardis ER, et al. Challenges of sequencing human genomes. *Brief Bioinform* 2010; 11:484-98
215. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 2007; 99:147-57
216. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* 2006; 103:5923-8
217. Teo YY. Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Curr Opin Lipidol* 2008; 19:133-43
218. Kitano H. Systems biology: a brief overview. *Science* 2002; 295:1662-4
219. Hudson TJ, Anderson W, Artz A, et al. International network of cancer genome projects. *Nature* 2010; 464:993-8

220. van Elmpt W, Ollers M, Dingemans AM, et al. Response Assessment Using 18F-FDG PET Early in the Course of Radiotherapy Correlates with Survival in Advanced-Stage Non-Small Cell Lung Cancer. *J Nucl Med* 2012;
221. Mayr NA, Huang Z, Wang JZ, et al. Characterizing Tumor Heterogeneity with Functional Imaging and Quantifying High-Risk Tumor Volume for Early Prediction of Treatment Outcome: Cervical Cancer as a Model. *International Journal of Radiation Oncology\*Biography\*Physics* 2012;
222. Digital Agenda for Europe [cited 2012 May 1]. Available from: [http://ec.europa.eu/information\\_society/digital-agenda/index\\_en.htm](http://ec.europa.eu/information_society/digital-agenda/index_en.htm)
223. Information Society - ICT for Health - home page [cited 2012 May 1]. Available from: [http://ec.europa.eu/information\\_society/activities/health/index\\_en.htm](http://ec.europa.eu/information_society/activities/health/index_en.htm),



# Chapter 3

## Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential PETCT imaging

Ruud van Stiphout, Guido Lammering, Jeroen Buijsen, Marco Janssen, Maria Antonietta Gambacorta, Pieter Slagmolen, Maarten Lambrecht, Domenico Rubello, Marcello Gava, Alessandro Giordano, Eric O. Postma, Karin Haustermans, Carlo Capirci, Vincenzo Valentini, Philippe Lambin

Published in: Radiotherapy and Oncology 2011; 98(1):126-33

## Abstract

### *Purpose*

To develop and validate an accurate predictive model and a nomogram for pathologic complete response (pCR) after chemoradiotherapy (CRT) for rectal cancer based on clinical and sequential PETCT data. Accurate prediction could enable more individualized surgical approaches, including less extensive resection or even a wait-and-see policy.

### *Methods and materials*

Population based databases from 953 patients were collected in four different institutes and divided into three groups: clinical factors (training: 677 patients, validation: 85 patients), pre-CRT PETCT (training: 114 patients, validation: 37 patients) and post-CRT PETCT (training: 107 patients, validation: 55 patients). A pCR was defined as ypT<sub>0</sub>N<sub>0</sub> reported by pathology after surgery. The data were analyzed using a linear multivariate classification model (support vector machine), and the model's performance was evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve.

### *Results*

The occurrence rate of pCR in the datasets was between 15% and 31%. The model based on clinical variables ( $AUC_{\text{train}}=0.61 \pm 0.03$ ,  $AUC_{\text{validation}}=0.69 \pm 0.08$ ) resulted in the following predictors: cT- and cN-stage, and tumor length. Addition of pre-CRT PET data did not result in a significantly higher performance ( $AUC_{\text{train}}=0.68 \pm 0.08$ ,  $AUC_{\text{validation}}=0.68 \pm 0.10$ ) and revealed maximal radioactive isotope uptake ( $SUV_{\text{max}}$ ) and tumor location as extra predictors. The best model achieved was based on the addition of post-CRT PET-data ( $AUC_{\text{train}}=0.83 \pm 0.05$ ,  $AUC_{\text{validation}}=0.86 \pm 0.05$ ) and included the following predictors: tumor length, post-CRT  $SUV_{\text{max}}$  and relative change of  $SUV_{\text{max}}$ . This model performed significantly better than the clinical model ( $p_{\text{train}} < .001$ ,  $p_{\text{validation}} = .056$ ).

### *Conclusions*

The model and the nomogram developed based on clinical and sequential PETCT data can accurately predict pCR, and can be used as a decision support tool for surgery after prospective validation.

## Introduction

Over the past decades, treatment outcomes for rectal cancer have changed dramatically. A better surgical technique, total mesorectal excision (TME), and the introduction of neoadjuvant treatments in locally advanced rectal cancer (LARC) have significantly decreased the risk of locoregional relapse.<sup>1,2</sup> In the last nine years at least seven published phase III trials have evaluated the role of adjuvant radiotherapy in rectal cancer.<sup>3</sup> These have provided an evidence base demonstrating the efficacy of both preoperative radiotherapy and preoperative concurrent chemotherapy (CRT). CRT has been reported to induce significant tumor downsizing and downstaging,<sup>4-6</sup> with a pathologic complete response (pCR) after CRT observed in 10% - 30% of patients.<sup>2,4-8</sup> Although some studies showed no correlation,<sup>9</sup> many others reported that patients showing a pCR following preoperative CRT have improved long-term outcomes including excellent local control rates and disease-free survival, regardless of their initial clinical T- and N-stages.<sup>10-13</sup>

However, despite the often phenomenal downsizing and sometimes even complete pathological responses after CRT, these patients are still operated with a standard extended surgical procedure due to the lack of reliable accurate preoperative diagnostic tools. However, it may be questioned whether a standard resection is still necessary, considering the good outcome of these patients reported with less invasive treatments.<sup>14,15</sup> If accurately selected, patients with a complete response (no residual tumor) may undergo a less extensive resection or even a so called 'wait-and-see' policy. Compared to standard surgery, the benefits of these treatments are reduced morbidity and mortality (e.g., anastomotic leakage, relaparotomy, wound and pelvic infection, abscess, colostomy, chronic wound healing disturbances, faecal or urinary incontinence and sexual dysfunction), improved quality of life and reduced treatment costs.

Thus, an accurate prediction of pCR can help in the selection of patients for more optimized treatment, sphincter-preserving surgery, less extensive resection, more intense radiation treatment, or even delayed surgery with a wait-and-see policy.<sup>2,3,16</sup> These considerations led to the overall goal of this study: to develop an accurate, data-driven model to predict pathologic complete response for rectal cancer patients as decision support for more individualized treatment approaches in the future.

The clinical variables associated with a better response to preoperative CRT include circumferential tumor extent, tumor differentiation, preoperative classification, carcinoembryonic antigen (CEA) level, distance from anal verge, and time to surgery.<sup>6,17,18</sup> Recently, it has also been suggested that PET imaging might be correlated

with tumor response after CRT in locally advanced rectal cancer. However, the studies involved used only a small number of patients, which meant that contradictory results were found. Further, only semi-quantitative PET measurements were used and analyzed with univariate statistics.<sup>4,5,7,19-26</sup> Multivariate analysis was performed in only one study, whose results lacked statistical significance.<sup>27</sup> Notably, no studies verified and validated their results with external datasets, despite the fact that this represents an important prerequisite for the generalizability of prediction models for other institutes.

In the current study, population based data from four different institutes were collected and used to train and validate predictive models for pCR. We hypothesized that the addition of PET imaging data to clinical variables significantly increases the performance of prediction models for pCR after CRT as compared to models based on clinical data alone.

The study was performed within the framework of a decision support system based on centralized datasets. The increasing amount of available patient information requires automatic methods for model building and analysis. Machine learning methods can be used to update the models continuously by feeding them with information of new patients. The increasing complexity of prediction models, too, means that the representation and interpretation of the results also become more important. Tools to enhance interpretation for the clinic include visualization techniques such as nomograms and graphical networks. Nomograms are statistical tools that enable users to calculate the overall probability of a specific clinical outcome for an individual patient.<sup>28</sup> In this study, the nomogram with the highest accuracy for the prediction of pCR is provided.

## Methods and materials

### *Study population*

Six population based datasets were collected in four institutes: Maastricht Clinic (GROW, MUMC, Maastricht, the Netherlands), Università Cattolica del S.Cuore (Rome, Italy), S. Maria della Misericordia Hospital (Rovigo, Italy) and University Hospital Gasthuisberg (Leuven, Belgium). In total, 953 patients met the criteria for inclusion: long-course RT with neoadjuvant chemotherapy and the availability of pathological outcome for pCR. Of these, 276 patients underwent a pre-CRT PET scan (one week before the start of CRT), and 169 patients had both pre- and post-CRT PET scans (one week before

surgery, and six to eight weeks after the end of CRT). The sequential PET data from Rovigo have already been published as a prospective study,<sup>20</sup> the Leuven data were collected prospectively for the BioCare project (LSHC-CT-2204-505785) and the rest of the data were gathered for a population-based study registered in the Dutch Trial Register (NTR2166). All compositions of the cohorts were approved by the local IRB committees. The patient characteristics are reported in Table 3.1. The datasets were divided into three groups, based on PET data availability: 1. clinical variables only, 2. clinical variables with pre-CRT PET variables (PET-pre), 3. clinical variables with both pre- and post-CRT PET variables (PET-post). For each group, a training set and an external validation set were defined. The training sets were used to identify the pCR predictors, while the validation sets were used to test the performance of the models in other centers. Datasets from a single center with the highest number of patients were used for training. A dataset was deemed not useful for external validation if it originated from the same center as the corresponding training set. The definition of the different combined training and validation sets is explained in Table 3.2, based on the datasets in Table 3.1.

The available clinical variables were age, gender (0: female, 1: male), clinical tumor (cT) and nodal (cN) stage, and two variables based on MRI (or endoscopy if MRI was unavailable): tumor location categorized in three levels (1: low, 0–5 cm from anal verge; 2: mid, 5–10 cm from anal verge; 3: high, >10 cm from anal verge) and tumor length (cm). For the patients who had PETCT scans, the tumors were semi-automatically contoured at Maastricht Clinic using dedicated software (TrueD, Siemens Medical, Erlangen, Germany). Standardized uptake-value (SUV) thresholding was based on the tumor-to-background signal ratio, with the gluteus muscle as reference background.<sup>29,30</sup> From the resulting tumor contour, maximal tumor diameter (MaxD), gross tumor volume (GTV), and maximal and mean SUV values within the GTV were calculated. If the post-CRT PETCT scan was available, the same variables were scored, and a response index (RI) for each variable was calculated. For variable  $X$ , the response index is the relative percent difference between the value of the post-CRT and pre-CRT and it was defined as  $RI = (X_{pre} - X_{post}) / X_{pre} \times 100\%$ . Thus, six variables were evaluated for the clinical dataset, 10 for the PET-pre dataset and 18 for the PET-post dataset. From these sets, the models selected subgroups of variables with significant predictive value for pCR.

All patients underwent surgery. Pathological complete response was defined as ypT<sub>0</sub>N<sub>0</sub>, extracted from the pathologic reports of surgical specimens. All other cases (ypT+ and/or ypN+) were considered non-responders, making the pCR a binary

outcome (0/1). The specimens were not re-evaluated centrally but the pathology protocols were very similar between institutes (3-5 mm slices of rectum tumor, intensified evaluation on several blocks of tissue at the tumor site, evaluation on 2-3 sublevels when no tumor tissue was found in initial block).

**Table 3.1** Patient characteristics for six datasets from four different institutes. Clinical, PET-pre and PET-post groups are defined. Percentages of the total patient numbers are given for binary or ordinal variables. Mean and standard deviation (SD) are given for continuous variables. x denotes missing values. RT = Radiotherapy, PF = per fraction.

Center	Maastricht		Rome		Rovigo	Leuven
Dataset	M1	M2	R1	R2	C1	L1
Period	2004–2006	2004–2006	1984–2008	2007–2008	2003–2007	2005–2007
# Patients	114	21	677	18	107	16
Clinical	Validation	-	Training	-	-	-
PET-pre	Training	-	-	Validation	-	Validation
PET-post	-	Validation	-	Validation	Training	Validation
Gender (%)						
Male	63	67	63	83	74	81
Female	37	33	37	17	26	19
Age (years)						
Mean	65.6	66.1	61.3	60.4	66.3	58.6
SD	10.0	10.6	10.2	7.1	10.8	10.1
cT-stage (%)						
1	0	0	0	0	0	0
2	1	0	3	11	0	0
3	68	81	86	56	90	94
4	30	14	11	33	10	6
x	1	5	0	0	0	0
cN-stage (%)						
0	25	38	23	17	51	0
1	48	48	45	33	38	62
2	26	10	30	50	10	38
x	1	4	2	0	1	0
cM-stage (%)						
0	73	71	100	94	100	100
1	25	19	0	6	0	0
x	2	10	0	0	0	0
ypT <sub>0</sub> N <sub>0</sub> (%)						
No	85	81	80	78	76	69
Yes	15	19	20	22	24	31
RT dose						
Mean	50.4	50.4	49.0	52.7	55.7	45.7
SD	0	0	5.5	3.3	3.1	1.8
RT dose PF (Gy)	1.8	1.8	1.8	1.8	2.2	1.8
Chemo Types (number)	1	1	11	2	1	1

### ***Statistical analysis***

Missing values in the dataset were substituted by the mean.<sup>31</sup> This method performed similarly to other, more complex substitution methods for small percentages of missing values (e.g., expectation-maximization imputation, regression estimation). No variables in the datasets exceeded 5% of missing values. Patients who missed tumor location and length in the clinical datasets (Roma: n = 132 and Maastricht: n = 29) were excluded because of too large amounts of missing data for these variables. All patient numbers stated in this paper were extracted after the missing value procedure. To compare the weights of significance assigned to the variables by the model, all variables were normalized by subtracting the mean, and then divided by the standard deviation.

To classify the complete responders and non-responders, a linear multivariate method suitable for binary classification from the machine learning field was used: the support vector machine (SVM).<sup>32</sup> The SVM variant used (proximal SVM or pSVM) performs equally accurately but much faster than normal support vector machines.<sup>33</sup> The different datasets' performances in predicting pCR were evaluated by analyzing the area under the curve (AUC) of the receiver operating characteristic (ROC) curve.<sup>34</sup> The maximum value of the AUC is 1.0, indicating a perfect prediction model; a value of 0.5 indicates a random chance of correct prediction.

To select the variables that contribute to pCR prediction, an exhaustive feature search was performed, with all possible variable combinations used as input for the pSVM model. The set of variables resulting in the highest AUC was selected as the final predictive set. To avoid over-fitting of the model through selection of the highest AUC, the variable sets resulting in AUCs that deviated less than 5% from the maximal AUC were compared to the final variable set. If conflicts occurred or if variables did not contribute significantly, selected variables were interchanged by considering their prevalence in the highly predictive sets, the factor analysis and the Spearman correlation coefficient (i.e., highly correlated and dependent variables are not present in the same predictive set). Furthermore, an extra univariate analysis was performed using the Wilcoxon rank sum test.

Classification methods normally require at least several hundred cases. Because of the relatively small number of available patients, two extra evaluation methods were used. The first was leave-one-out (LOO) cross-validation, used to calculate an AUC for the training set. In LOO cross-validation, a single patient is selected from the original training dataset and used as the validation dataset, while the data from the remaining

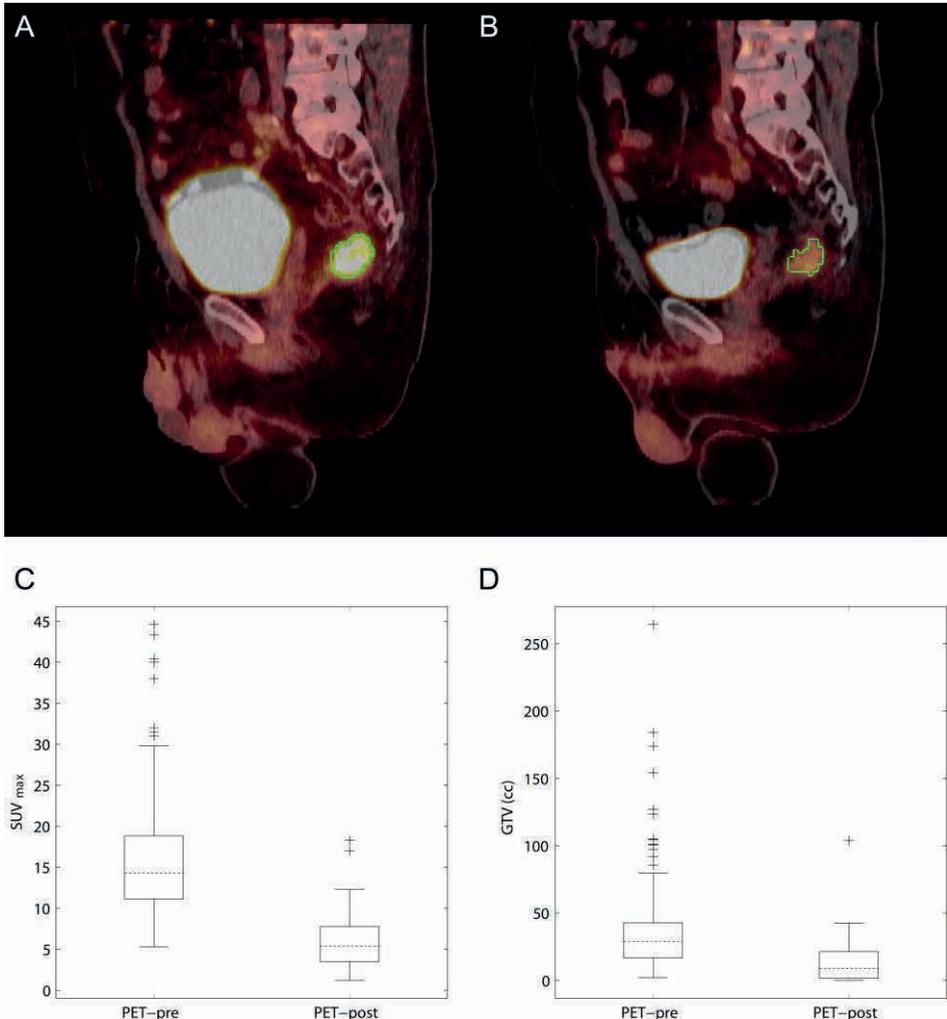
patients are used to train the model. This is repeated until all patients have been selected once for validation. However, no LOO cross-validation was used for the external dataset. The second evaluation method was bootstrapping, which results in a more accurate approximation of the real dataset distribution.<sup>35</sup> This means that 1000 datasets are generated from the original dataset containing  $n$  patients by selecting these  $n$  patients, but with resampling (i.e., patients can be present in the dataset more than once). For every bootstrapped dataset, an AUC was calculated. The mean AUC with the corresponding standard deviation was then calculated with size 1000. This non-parametric method allows comparison of the confidence intervals of the AUCs of different datasets without making assumptions about the AUC distributions.<sup>36</sup> The distribution of the difference in mean AUC ( $\Delta$ AUC) between the datasets was tested by calculating the two-sided p-value, i.e., the fraction of  $\Delta$ AUC samples smaller or larger than zero (depending on the dominant sign of  $\Delta$ AUC).

Nomograms can reduce statistical predictive models to a single numerical estimate of the probability of an event, and visualize the effect of each selected variable on this probability.<sup>37</sup> The model output of the pSVM models consists of assigned weights for each variable and an offset. The probability of a patient having a pCR can be calculated using logistic regression on the pSVM output.<sup>38</sup> The complete procedure to convert SVM output to a nomogram is described in detail elsewhere.<sup>39</sup> Developing a nomogram requires threshold selection in the ROC curve. For response prediction specificity is most important, because it is not preferred to predict non-responders as responders, which would result in under-treatment. Therefore, the threshold was selected in such a way that at least 90% of non-responders were correctly predicted. Partial ROC curve optimization has been tested but it had no gain for specificity compared to overall AUC maximization.<sup>40</sup> Calibration of the nomogram, i.e., the agreement between predicted probability of complete response and true probability in the population, was performed by an assessment of the overall agreement and the Hosmer-Lemeshow statistic in four subgroups of patients in the validation data. The nomogram algorithm was implemented in MATLAB (version 7.1, MathWorks Inc., Natick, MA), as were all algorithms described in this section.

## Results

The occurrence of pCR in the patient population varied between 15% and 31% (mean: 21.8%, SD: 5.4%) depending on the dataset (Table 3.1). A first evaluation of CRT's effect on the tumor demonstrated significant downsizing of the tumor in the PETCT,

and a significant decrease in metabolic activity within the tumor (Figure 3.1). Both gross tumor volume and maximal SUV decreased significantly between the pre- and post-CRT PETCT scans ( $p < .001$ ).



**Figure 3.1** (A) Tumour contour in a fused FDG-PETCT made pre-CRT. (B) Corresponding post-CRT FDG-PETCT scan with tumour contour. (C) Boxplot of  $SUV_{max}$  on PET-scans made pre-CRT and post-CRT; significant decrease:  $p < .001$  (D) Boxplot of the GTV for the case of pre-CRT and post-CRT; significant decrease:  $p < .001$ .

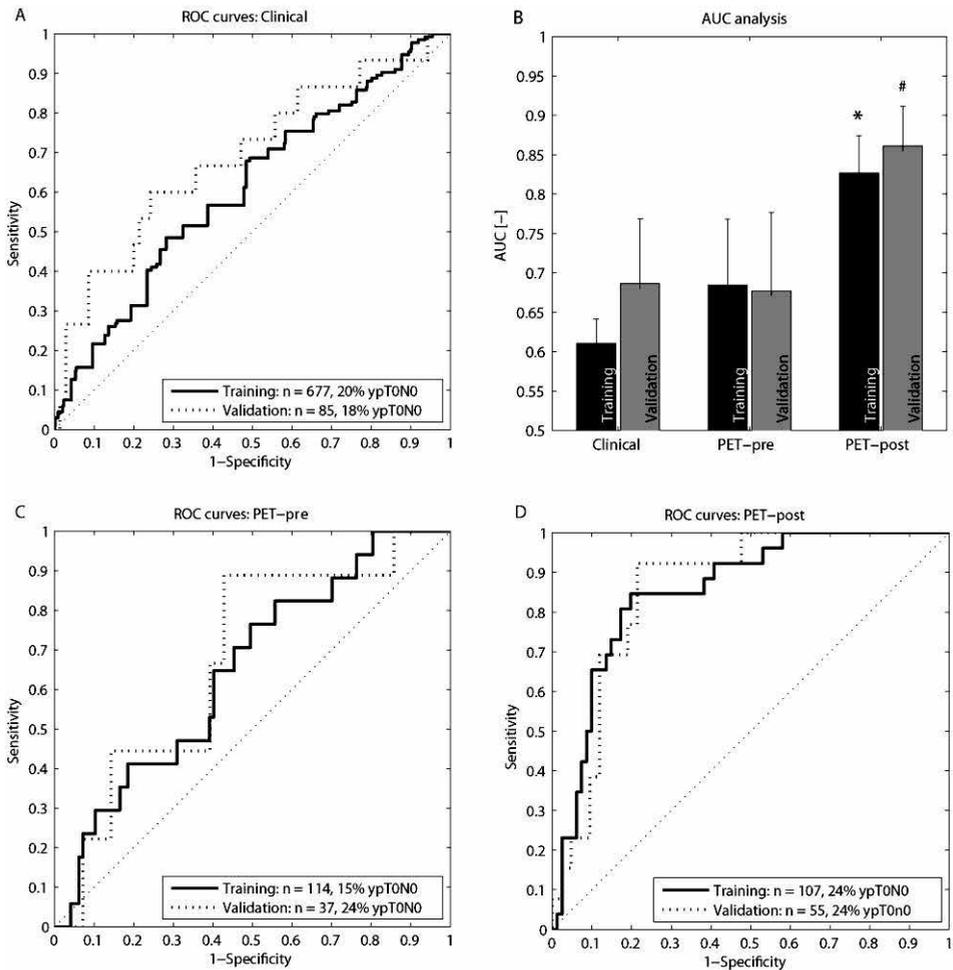
Table 3.2 shows the predictor selection results and the ROC curve analysis. For the clinical dataset, the univariate analysis reveals three variables significantly associated with pCR (95% confidence interval): tumor length ( $p < .001$ ), cN-stage ( $p = .001$ ), and cT-

stage ( $p=.001$ ). These variables were also selected in the multivariate analysis. The normalized weights assigned to them by the pSVM model are tumor length (-0.085), cT-stage (-0.074), and cN-stage (-0.060). The selected variables were ranked in importance (i.e., weights). The sign of the weights can be interpreted by the effect on the probability of a pCR. For a negative sign, this probability decreases when the variable increases. For the clinical dataset, this means that the probability of a pCR increases for small tumor lengths and low cT- and cN-stages. The predictive performance of the clinical dataset for pCR, expressed by the AUC of the ROC curve, is  $0.61 \pm 0.03$  (mean  $\pm$  SD) for the training set and  $0.69 \pm 0.08$  for the external validation set.

**Table 3.2** Predictor selection and ROC analysis. Predictive variables are given with their corresponding assigned normalized weights from multivariate analysis (MVA). For each variable the p-value from univariate analysis (UVA) is given. Mean AUC and standard deviation (SD) are given for each variable set. RI = response index, SUV = standard uptake value, MaxD = maximal diameter (PETCT).

Variable set	Type	Size	Predictors (MVA)	Weights (MVA)	p-value (UVA)	AUC	SD
Clinical	Training (R1)	677	Tumor length	-0.085	<.001	0.61	0.03
			cT-stage	-0.074	.001		
			cN-stage	-0.060	.001		
	Validation (M1)	85	-	-	-	0.69	0.08
Clinical + PET-pre	Training (M1)	114	MaxD <sub>pre</sub>	-0.12	.003	0.68	0.08
			cN-stage	-0.12	.001		
			Tumor location	0.094	.84		
			SUV <sub>max-pre</sub>	-0.087	.29		
		Validation (R2, L1)	34	-	-	-	0.68
Clinical + PET-pre + PET-post	Training (C)	107	RI <sub>SUVmax</sub>	0.20	<.001	0.83	0.05
			Tumor length	-0.20	<.001		
			SUV <sub>max-post</sub>	-0.14	<.001		
		Validation (M2, R2, L1)	55	-	-	-	0.86

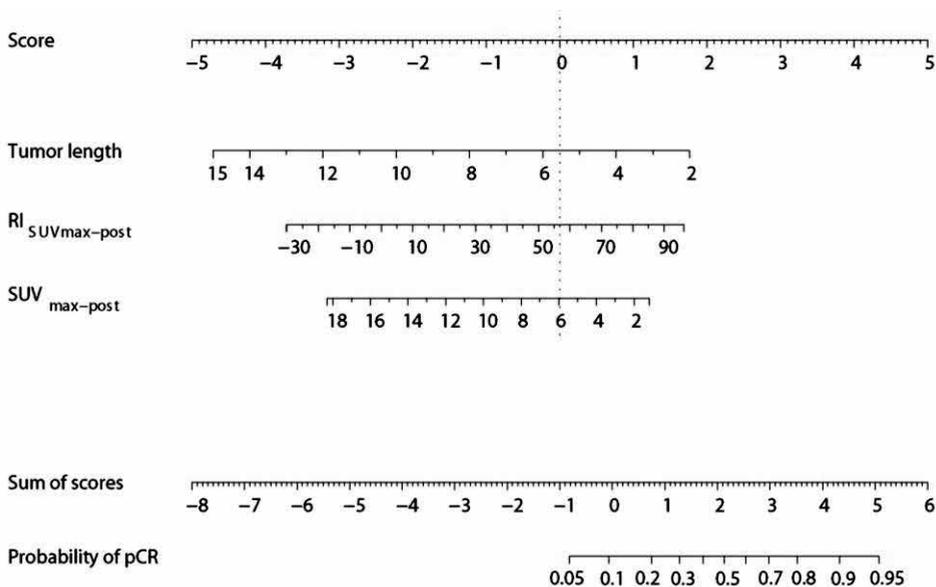
For the dataset with pre-CRT PET data, the multivariate analysis selected these variables (ranked by weight): maximal diameter (-0.12), cN-stage (-0.12), tumor location (0.094), and SUV<sub>max</sub> (-0.087). This resulted in a high probability of pCR for patients with small maximal tumor diameters, low cN-stage, high tumor locations, and small maximal metabolic activity. Maximal diameter ( $p=.003$ ) and cN-stage ( $p=.001$ ) were selected by univariate analysis, while the other two variables were not.



**Figure 3.2** ROC curves of training and validation datasets for the clinical set (A), the PET pre-CRT set (C) and the PET post-CRT set (D). The straight dashed line represents a random prediction model. The bar plot (B) shows the corresponding mean AUC for each dataset and the standard deviation (error bars). There was a significant difference with clinical datasets of (\*)  $p < .05$  and (#)  $p < .06$ .

The AUCs for the training and validation set were both 0.68, but the SD differed (0.08 and 0.10 respectively). The dataset including the post-CRT PET data resulted in the highest performance:  $AUC_{\text{train}} = 0.83 \pm 0.05$  and  $AUC_{\text{validation}} = 0.86 \pm 0.05$ . The response index for  $SUV_{\text{max}}$  (0.20), tumor length (-0.20), and the post-CRT  $SUV_{\text{max}}$  were found to be predictive for pCR and significantly associated with pCR in the univariate analysis ( $p < .001$ ).

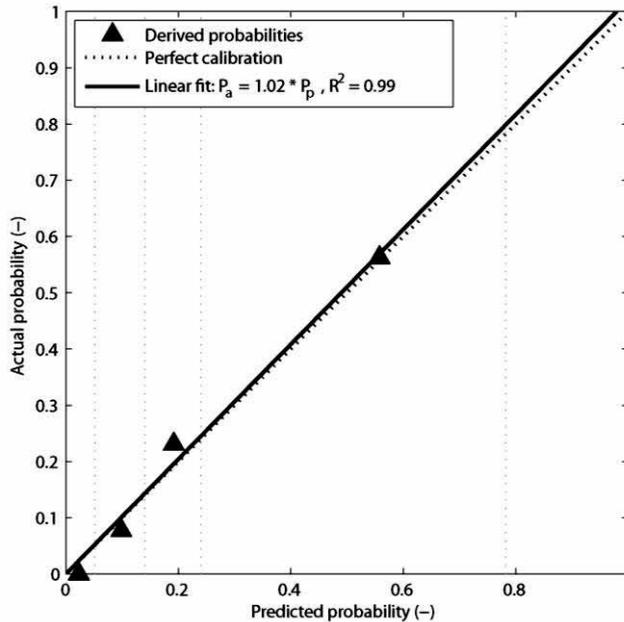
In evaluating the predictive value of the additional PET data to the clinical data, only the AUCs of the post-CRT PET data differed significantly from the clinical dataset AUC (Figure 3.2). The p-value for the AUC difference for the training set was  $<0.001$ , while that for the validation sets was 0.056 (just outside the 95% confidence interval). When only post-CRT PET data were used for the models (i.e., no clinical variables), the significant difference between the AUCs and the clinical dataset was no longer observed (training:  $p=.47$ , validation:  $p=.58$ ). This indicated that a combination of both clinical and PET data was required to reach a significantly higher performance when using PET as a predictive imaging modality.



**Figure 3.3** Nomogram for PET post-CRT dataset. A score for each predictor can be read out at the top scale (Score). All summed scores (Sum of scores scale) can be converted directly to the probability of responding with a pCR ( $ypT_0N_0$ ). The probability scale is the only logarithmic scale.

The assigned weights for all the predictors formed the basis for the construction of the nomogram. The nomogram based on the post-CRT dataset is provided in Figure 3.3. The nomogram performs with a sensitivity of 0.62 and a specificity of 0.88 for the validation data. In the training phase these were respectively 0.65 and 0.90. The calibration of the nomogram (Figure 3.4) with the validation data reveals that the overall predicted and the actual probability are equal (23.6%,  $OR=1.0$ ). If the validation data are divided into four equally numbered groups, the Hosmer-Lemeshow test results in a p-value of 0.78, which means a good calibration in this test ( $p>.05$ ). The

linear fit through these probabilities results in a slope of 1.02 with  $R^2$  of 0.99, confirming a good balance between calibration and discrimination.



**Figure 3.4** Calibration of the nomogram for the validation data. For the four equally numbered subgroups (vertical lined intervals in figure), the predicted probability of a pCR and the actual fraction in the population were evaluated. The dashed line represents perfect calibration and the solid line is the linear fit of the calibration data.  $P_a$ : actual probability,  $P_p$ : predicted probability,  $R^2$ : coefficient of determination

## Discussion

We have developed predictive models based on clinical and PET-based data for pathologic complete response in patients diagnosed with rectal cancer. The performance of these models was externally validated using patient cohorts from different institutes treated with long-course preoperative chemoradiotherapy. The models showed that the accuracy of the predictions increased over time, i.e., when more information became available. Information from PETCT scans significantly improved the performance of the models.

The significant difference in AUCs that we reported between the performance of the clinical model and the post-CRT PET data model reflects what others have found in

their post-treatment PET analyses; like us, some have reported (significant) indications that the response index and post-treatment  $SUV_{max}$  are predictive for response, while the pre-treatment PET data do not provide enough predictive power.<sup>12,19,27</sup> However, our PET-based models also contain clinical variables, which appeared to be necessary to obtain the high performance provided in Table 3.2. The most important clinical variables were tumor length and maximal diameter, which were selected in the models and are significantly correlated (spearman  $\rho = 0.55$ ,  $p < .001$ ). Overall, this means that the dominant tumor dimension in combination with (differences in) the maximal metabolic activity inside the tumor is the most predictive variable set for pCR, which was confirmed in the external datasets.

Whether the corresponding AUC of 0.86 is accurate enough for clinical practice depends on the choice of the threshold in the ROC curve. A high specificity is preferred over a high sensitivity to avoid possible under-treatment (less surgery when surgery is required) rather than over-treatment (standard treatment when less surgery could have been considered). The provided nomogram focuses on specificity (training: 0.90, validation 0.88). Selecting higher specificities results in fast decreasing sensitivities. Careful follow-up is therefore necessary for the patients selected for a 'wait-and-see' policy to detect any possible local recurrences early on. To gain more specificity in the future, the addition of new variables and the other classification methods would have to be considered.

The nomogram performs well, i.e., the distribution of the probability of a pCR provided by the nomogram represents the true distribution in the data, confirmed by overall calibration, calibration of the slope and Hosmer-Lemeshow test (Figure 3.4). Because of the number of events and the division of the patient cohorts into few probability intervals, the higher probabilities occur much less frequently and are thus the least accurate. Therefore, prospective validation of the model and the nomogram is required to ensure sufficient statistical power for clinical application of the models. Besides the number of patients to increase the models' accuracy, more predictors could be added to increase the models' performance, including biological variables such as gene signatures<sup>41</sup> and blood biomarkers, and also more imaging variables from (perfusion) CT and (diffusion) MRI. The first indications have also appeared that PETCT data during CRT may be highly predictive for response.<sup>25,26,42</sup> This time point is more favorable than post-CRT because of the possibility of earlier treatment changes and the decreased presence of inflammatory rectum cases, potentially causing impaired evaluation of fused PETCT scans. After prospective validation of the model, an intervention trial with less surgery for patients with a high probability for pCR will be performed.

The population based collected datasets date back five years, except for the clinical Roma database, which was collected from 1984 onward. Therefore, this dataset shows a higher variety in treatment schemes than the other datasets. This could explain the discrepancy of the higher prediction performance of the clinical validation set. On the other hand, the validation set is much smaller, implying that the distribution of data could not be representative of the true distribution. The consequence of population based data collection is that treatment protocols are not well tuned. This results in, for example, small differences in irradiation schemes and deviations in the evaluation of pathology outcome. Ideally, pathology is reviewed centrally to reduce the intra- and inter-observer variabilities for the outcome measure. However, in this study the quality of pathology is acceptable because of the prospective nature of most datasets and because the outcome was limited to only complete response evaluation. Also, glucose correction for SUV values was not applied to all datasets. However, minor variation in treatment schemes can be seen as an advantage because it leads to higher generalizability for other centers. In other words, the model still performs well, despite the disparities mentioned here.

In conclusion, we have shown that sequential PETCT data in combination with clinical variables significantly increase the performance of prediction models for pathologic complete response. So far, this is the largest study of its kind and the only one that used external datasets for validation. The dominant tumor dimension and the maximal uptake of radioactive isotopes in the tumor as well as its relative difference between PET scans were found to be the best predictors for pCR resulting in very good overall performance AUC's of 0.83 and 0.86 for training and validation, respectively. Including also biological and other imaging variables will probably further improve the performance. When prospectively validated, the model and the nomogram therefore provide a valuable decision support for more individualized treatment approaches in the future.

Note: The predictive models in this paper are published on the website <http://www.predictcancer.org>

## References

1. Adjuvant radiotherapy for rectal cancer: a systematic overview of 8,507 patients from 22 randomised trials. *Lancet* 2001; 358:1291-304
2. Valentini V, Beets-Tan R, Borras JM, et al. Evidence and research in rectal cancer. *Radiother Oncol* 2008; 87:449-74
3. Valentini V, Aristei C, Glimelius B, et al. Multidisciplinary Rectal Cancer Management: 2nd European Rectal Cancer Consensus Conference (EURECA-CC2). *Radiother Oncol* 2009; 92:148-63
4. Capirci C, Rampin L, Erba PA, et al. Sequential FDG-PET/CT reliably predicts response of locally advanced rectal cancer to neo-adjuvant chemo-radiation therapy. *Eur J Nucl Med Mol Imaging* 2007; 34:1583-93
5. Capirci C, Rubello D, Chierichetti F, et al. Long-term prognostic value of 18F-FDG PET in patients with locally advanced rectal cancer previously treated with neoadjuvant radiochemotherapy. *AJR Am J Roentgenol* 2006; 187:W202-8
6. Valentini V, Coco C, Cellini N, et al. Ten years of preoperative chemoradiation for extraperitoneal T3 rectal cancer: acute toxicity, tumor response, and sphincter preservation in three consecutive studies. *Int J Radiat Oncol Biol Phys* 2001; 51:371-83
7. Vliegen RF, Beets-Tan RG, Vanhauten B, et al. Can an FDG-PET/CT Predict Tumor Clearance of the Mesorectal Fascia after Preoperative Chemoradiation of Locally Advanced Rectal Cancer? *Strahlenther Onkol* 2008; 184:457-464
8. O'Neil BH, Tepper JE. Current options for the management of rectal cancer. *Curr Treat Options Oncol* 2007; 8:331-8
9. Pucciarelli S, Toppan P, Friso ML, et al. Complete pathologic response following preoperative chemoradiation therapy for middle to lower rectal cancer is not a prognostic factor for a better outcome. *Dis Colon Rectum* 2004; 47:1798-807
10. Vecchio FM, Valentini V, Minsky BD, et al. The relationship of pathologic tumor regression grade (TRG) and outcomes after preoperative therapy in rectal cancer. *Int J Radiat Oncol Biol Phys* 2005; 62:752-60
11. Habr-Gama A, Perez RO, Nadalin W, et al. Long-term results of preoperative chemoradiation for distal rectal cancer correlation between final stage and survival. *J Gastrointest Surg* 2005; 9:90-9; discussion 99-101
12. Capirci C, Valentini V, Cionini L, et al. Prognostic value of pathologic complete response after neoadjuvant therapy in locally advanced rectal cancer: long-term analysis of 566 ypCR patients. *Int J Radiat Oncol Biol Phys* 2008; 72:99-107
13. Rodel C, Martus P, Papadopoulos T, et al. Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. *J Clin Oncol* 2005; 23:8688-96
14. Borschitz T, Wachtlin D, Mohler M, et al. Neoadjuvant chemoradiation and local excision for T2-3 rectal cancer. *Ann Surg Oncol* 2008; 15:712-20
15. Lezoche G, Baldarelli M, Guerrieri M, et al. A prospective randomized study with a 5-year minimum follow-up evaluation of transanal endoscopic microsurgery versus laparoscopic total mesorectal excision after neoadjuvant therapy. *Surg Endosc* 2008; 22:352-8
16. Habr-Gama A, Perez RO, Nadalin W, et al. Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results. *Ann Surg* 2004; 240:711-7; discussion 717-8

17. Berger C, de Muret A, Garaud P, et al. Preoperative radiotherapy (RT) for rectal cancer: predictive factors of tumor downstaging and residual tumor cell density (RTCD): prognostic implications. *Int J Radiat Oncol Biol Phys* 1997; 37:619-27
18. Das P, Skibber JM, Rodriguez-Bigas MA, et al. Predictors of tumor response and downstaging in patients who receive preoperative chemoradiation for rectal cancer. *Cancer* 2007; 109:1750-5
19. Amthauer H, Denecke T, Rau B, et al. Response prediction by FDG-PET after neoadjuvant radiochemotherapy and combined regional hyperthermia of rectal cancer: correlation with endorectal ultrasound and histopathology. *Eur J Nucl Med Mol Imaging* 2004; 31:811-9
20. Capirci C, Rubello D, Pasini F, et al. The Role of Dual-Time Combined 18-Fluorideoxyglucose Positron Emission Tomography and Computed Tomography in the Staging and Restaging Workup of Locally Advanced Rectal Cancer, Treated with Preoperative Chemoradiation Therapy and Radical Surgery. *Int J Radiat Oncol Biol Phys* 2009;
21. Denecke T, Rau B, Hoffmann KT, et al. Comparison of CT, MRI and FDG-PET in response prediction of patients with locally advanced rectal cancer after multimodal preoperative therapy: is there a benefit in using functional imaging? *Eur Radiol* 2005; 15:1658-66
22. Guillem JG, Moore HG, Akhurst T, et al. Sequential preoperative fluorodeoxyglucose-positron emission tomography assessment of response to preoperative chemoradiation: a means for determining longterm outcomes of rectal cancer. *J Am Coll Surg* 2004; 199:1-7
23. Kalff V, Duong C, Drummond EG, et al. Findings on 18F-FDG PET scans after neoadjuvant chemoradiation provides prognostic stratification in patients with locally advanced rectal carcinoma subsequently treated by radical surgery. *J Nucl Med* 2006; 47:14-22
24. Melton GB, Lavelly WC, Jacene HA, et al. Efficacy of preoperative combined 18-fluorodeoxyglucose positron emission tomography and computed tomography for assessing primary rectal cancer response to neoadjuvant therapy. *J Gastrointest Surg* 2007; 11:961-9; discussion 969
25. Rosenberg R, Herrmann K, Gertler R, et al. The predictive value of metabolic response to preoperative radiochemotherapy in locally advanced rectal cancer measured by PET/CT. *Int J Colorectal Dis* 2009; 24:191-200
26. Janssen MH, Ollers MC, Riedl RG, et al. Accurate prediction of pathological rectal tumor response after 2 weeks of pre-operative radiochemotherapy using FDG-PET-CT imaging. *Int J Radiat Oncol Biol Phys* 2009;
27. Konski A, Li T, Sigurdson E, et al. Use of molecular imaging to predict clinical outcome in patients with rectal cancer after preoperative chemotherapy and radiation. *Int J Radiat Oncol Biol Phys* 2009; 74:55-9
28. Eastham JA, Kattan MW, Scardino PT. Nomograms as predictive models. *Semin Urol Oncol* 2002; 20:108-15
29. Daisne JF, Sibomana M, Bol A, et al. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiother Oncol* 2003; 69:247-50
30. Ollers M, Bosmans G, van Baardwijk A, et al. The integration of PET-CT scans from different hospitals into radiotherapy treatment planning. *Radiother Oncol* 2008; 87:142-6
31. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods* 2002; 7:147-77
32. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2007; 2:59-77
33. Fung GM, Mangasarian OL: Proximal support vector machine classifiers, International Conference on Knowledge Discovery and Data Mining. San Francisco (California, USA), ACM, 2001, pp 77-86

34. Pepe MS. Receiver Operating Characteristic Methodology. *Journal of the American Statistical Association* 2000; 95
35. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med* 2000; 19:1141-64
36. Niang N, Saporta G: Resampling ROC curves, *Statistics for Data Mining, Learning and Knowledge Extraction*. Aveiro (Portugal), 2007
37. Shariat SF, Capitanio U, Jeldres C, et al. Can nomograms be superior to other prediction tools? *BJU Int* 2009; 103:492-5; discussion 495-7
38. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008; 26:1364-70
39. Jakulin A, Možina M, Demšar J, et al: Nomograms for visualizing support vector machines, *International Conference on Knowledge Discovery and Data Mining*. Chicago (Illinois, USA), ACM, 2005, pp 108 - 117
40. Dodd LE, Pepe MS. Partial AUC estimation and regression. *Biometrics* 2003; 59:614-23
41. Starmans MH, Krishnapuram B, Steck H, et al. Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. *Br J Cancer* 2008; 99:1884-90
42. Cascini GL, Avallone A, Delrio P, et al. 18F-FDG PET is an early predictor of pathologic tumor response to preoperative radiochemotherapy in locally advanced rectal cancer. *J Nucl Med* 2006; 47:1241-8

# Chapter 4

Nomogram predicting response after chemo-radiotherapy in rectal cancer using sequential PETCT imaging: a prospective study with external validation

Ruud van Stiphout, Vincenzo Valentini, Jeroen Buijsen, Guido Lammering, Elisa Meldolesi, Maria Antonietta Gambacorta, Andre Dekker, Philippe Lambin

Submitted

## Abstract

### *Purpose*

To develop and externally validate a predictive model for pathologic complete response (pCR) for locally advanced rectal cancer (LARC) based on clinical and early sequential  $^{18}\text{F}$ -FDG PETCT imaging.

### *Methods and materials*

Prospective data (including the THUNDER trial) was used to train (N=112, Maastricht clinic) and validate (N=78, Gemelli Hospital) the model for pCR (ypT<sub>0</sub>N<sub>0</sub>). All patients received long-term chemoradiotherapy (CRT) and surgery. Clinical parameters were age, gender, clinical tumor (cT) stage and clinical nodal (cN) stage. PET parameters were  $\text{SUV}_{\text{max}}$ ,  $\text{SUV}_{\text{mean}}$ , metabolic tumor volume (MTV) and maximal tumor diameter, for which response indices between pre-treatment and intermediate scan were calculated. Using multivariate logistic regression, three probability groups for pCR were defined.

### *Results*

The pCR rates were 21.4% (training) and 23.1% (validation). The selected predictive features for pCR were cT-stage, cN-stage, response index of  $\text{SUV}_{\text{mean}}$  and maximal tumor diameter during treatment. The models' performances (AUC) were 0.78 (training) and 0.70 (validation). The high probability group for pCR resulted in 100% correct predictions for training and 67% for validation.

### *Conclusions*

The developed predictive model for pCR is accurate and externally validated. This model may assist in treatment decisions during CRT to select complete responders for a wait-and-see policy, good responders for extra RT boost and bad responders for additional chemotherapy.

## Introduction

An early prediction of pathologic complete response (pCR) for a locally advanced rectal cancer (LARC) patient is valuable because it allows for individualized treatment reorientation.<sup>1,2</sup> The standard treatment for LARC patients is preoperative chemoradiotherapy (CRT) followed by surgery. The neo-adjuvant treatment, intended to control pelvic disease and improve the chance of sphincter preservation, results in a pathological complete response (pCR) in 15-30% of the patients.<sup>3,4</sup> For these complete responders a wait-and-see policy after CRT is a possibility in order to reduce treatment-related morbidity and mortality, for which excellent results are reported.<sup>5</sup> This decision requires however a very accurate prediction and assessment of complete response. Other treatment options under consideration are a radiotherapy boost after CRT for good responding patients to achieve more pCRs<sup>6</sup> and additional chemotherapy administration after CRT for the worst responding patients.<sup>7</sup> Both these options require an early assessment of response even during CRT. Currently the leading candidate predictive marker for histopathological response prediction in LARC is <sup>18</sup>F-fluorodeoxyglucose (FDG) positron emission tomography (PET) imaging. A meta-analysis from 2012 confirmed the added value of PET imaging, especially for intermediate PET imaging (during CRT).<sup>8</sup> However, most studies evaluated pre-CRT versus post-CRT PET imaging. Besides that an early prediction is preferred for treatment reorientation, later predictions may also be affected by CRT-induced inflammatory tissue, which presents tumor equivalent signal on FDG-PET scans.<sup>9</sup> This recognition resulted in more early response assessment studies in the last few years (Table 4.1). The limitations of these studies were their small sample sizes (N=20-42), the main focus on good versus bad responders (not pCR), the univariate setting in which analyses were performed and the lack of validation. To increase the clinical applicability of these decision making tools, they need to be based on more evidence (i.e. larger number of patients and external validation), be trained on several data sources<sup>10</sup> and they require focus on outcomes that are more relevant in terms of decisions, like pCR for a possible wait-and-see policy. We hypothesize that models with these requirements are the most suitable for decision making in clinical practice. The aim of this study is therefore to develop an externally validated multivariate predictive model for pCR combining clinical, pre-treatment and intermediate FDG-PETCT imaging parameters based on a prospective study. After development of a nomogram and the evaluation of its accuracy, risk group definition based on these predictions may provide decision support to clinicians for LARC patients (Figure 4.1).

Table 4.1 All studies for response prediction in locally advanced rectal cancer using early sequential PET imaging

Study	Year	Accrual time	N	RT dose (Gy)	Chemo	2 <sup>nd</sup> PET (days)	Surg (weeks)	Outcome	PA	Predictors	AUC
Avallone <sup>20</sup>	2012	NR	42	45	5-FU LFU Oxa Ral	12	8	Resp	TRG	SUV <sub>I</sub> mean SUV <sub>I</sub> max ΔSUV <sub>I</sub> mean ΔSUV <sub>I</sub> max	NR
Goldberg <sup>22</sup>	2012	2008.2-2009.7	20	50.4-55	5-FU Cap UFT	8	6-10	pCR Resp	TRG	pCR: ΔSUV <sub>I</sub> max resp: none	NR
Leibold <sup>24</sup>	2011	2001-2005	27	50.4	5-FU	8-14	5-8	Resp	95% <sup>a</sup>	VRS	NR
Lambrecht <sup>27</sup>	2010	2005.5-2009.8	22	50.4	5-FU	14-16	6-8	pCR	TNM	ΔSUV <sub>I</sub> max	NR
Janssen <sup>11b</sup>	2009	2007.4-2009.3	30	50.4	Cap	8+15 <sup>c</sup>	6-8	Resp	TRG	ΔSUV <sub>I</sub> max ΔSUV <sub>I</sub> mean	0.87
Guerra <sup>23</sup>	2009	2006.5-2009.3	28	50.4	5-FU	21	12	Resp	TRG	SUV <sub>I</sub> mean	0.83
Rosenberg <sup>30</sup>	2008	2006.3-2007.1	30	45	5-FU	14	5	Resp	TRG <sup>d</sup>	None <sup>e</sup>	0.70
Cascini <sup>21</sup>	2006	NR	33	45	5-FU LFU Oxa Ral	12	8	Resp	TRG	SUV <sub>I</sub> max SUV <sub>I</sub> mean ΔSUV <sub>I</sub> mean ΔSUV <sub>I</sub> max	NR

2<sup>nd</sup> PET: time between start of chemoradiotherapy and the intermediate PET-scan, 5-FU: fluorouracil, Cap: capecitabine, ΔSUV: (relative) difference between pretreatment and intermediate standard uptake value of the PET contour, LFA: levofolnic acid, N: number of patients in study, NR: not reported, Oxa: oxaliplatin, pCR: pathological complete response, Ral: raltitrexed, Resp: tumor response, RT: radiotherapy, Surg: time between end of chemoradiotherapy and surgery, SUV<sub>I</sub>: intermediate standard uptake value of the PET contour, VRS: visual response score, UFT: tegafur-uracil

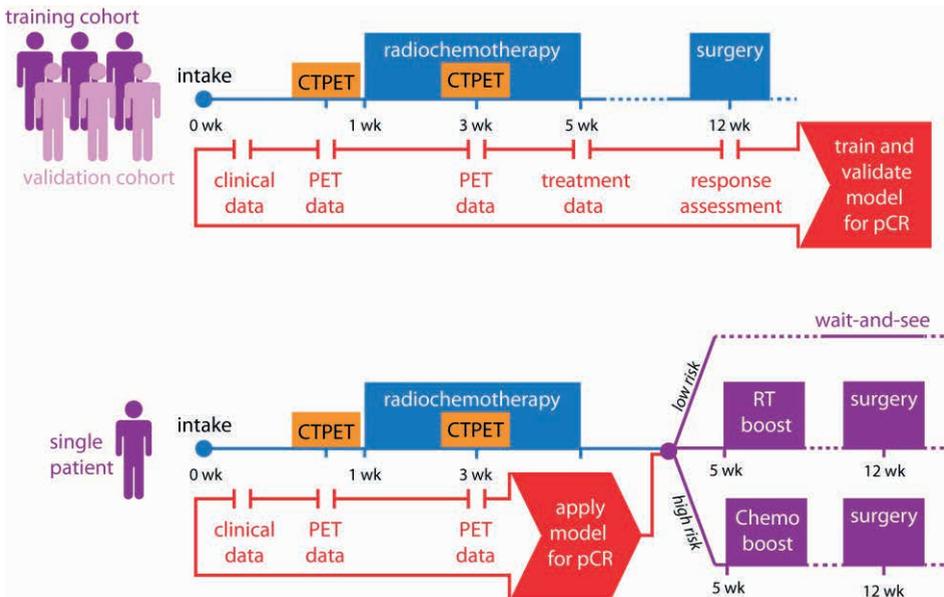
<sup>a</sup> Pathologic response >95%

<sup>b</sup> The separate validation study is not reported<sup>12</sup>.

<sup>c</sup> Results reported for day 15 (optimal time point in this study)

<sup>d</sup> According to Becker standard, other studies used Mandard standard

<sup>e</sup> Trend for ΔSUV<sub>I</sub>mean (p=0.085)



**Figure 4.1** Schematic overview of prediction model development (top) and the proposed application of the model in clinical practice after it has been tested in a clinical trial with a control arm (bottom).

## Methods and materials

### Study population

All data were prospectively collected (with written informed consent) between January 2007 and March 2012 within two institutes: Maastricht Clinic (GROW, MUMC, Maastricht, the Netherlands) and Università Cattolica del S.Cuore (Rome, Italy). The following prospective observational studies were involved: a study (2007-2009) involving 47 patients from Maastricht,<sup>11,12</sup> a pilot study (2007-2009) with 19 patients from Rome and a multicentre study (2009-2012) involving one protocol for both institutes (Maastricht: 65 patients, Rome: 59 patients) with acronym THUNDER (Theragnostic Utilities for Neoplastic DisEases of the Rectum, NCT00969657). All patients from Maastricht were pooled and used to train a prediction model for pCR (N=112). The pooled datasets from Rome were used for external validation of the model (N=78). The study inclusion criteria were: histological proven rectal cancer (primary tumors), UICC stage I-III, no recurrences, only concurrent chemoradiotherapy treatment, minimal age of 18 years, and no previous radiotherapy to the pelvis. The

available clinical variables used as candidate prognostic and predictive factors were age, gender, clinical tumor (cT) and nodal (cN) stage. The criteria followed to consider tumor nodal involvement at MRI were related to border contour (sharply demarcated or irregular border) and signal intensity characteristics (homogeneous or inhomogeneous) or size >8 mm.<sup>13,14</sup> All patients from Maastricht were treated preoperatively with radiotherapy (28 fractions of 1.8 Gy, 5 fractions/week) and concomitant chemotherapy (capecitabine, 825 mg/m<sup>2</sup>, twice daily), followed by a total mesorectal excision 6-8 weeks after the end of CRT. A minority of the thunder patients (N=11) with a clinical complete response (assessed using post-CRT MRI and endoscopy) were enrolled in a parallel study where a surgical wait-and-see policy was applied.<sup>5</sup> Some patients from Rome were also treated with 50.4Gy schedule, but 78.2% of the patients were treated with 25x1.8Gy schedule and a RT boost of 10Gy. The majority of the Rome patients (N=62) received a combination of capecitabine (1300 mg/m<sup>2</sup> daily) and oxaliplatin (60 mg/m<sup>2</sup> once a week for 5 weeks with 55.0Gy RT or 130 mg/m<sup>2</sup> at 3 time points with 50.4Gy RT), and the others capecitabine only (1650 mg/m<sup>2</sup> daily with 50.4Gy RT or 1300 mg/m<sup>2</sup> daily with 55 Gy RT, N=14) or raltitrexed (3 mg/m<sup>2</sup> at 3 time points, N=2).

### ***PETCT imaging***

All patients underwent a pre-CRT PET scan (one week before the start of CRT) and an intermediate PET scan (two weeks after the start of CRT). All Maastricht PETCT scans were performed by use of a dedicated Siemens Biograph 40 TruePoint PETCT simulator (Siemens Medical, Erlangen, Germany). Rome scans were performed with a 3D GEMINI GXL PETCT scanner with 16 channels CT (Philips Healthcare, Cleveland, OH). The PET acquisition settings were reported before and were calibrated for both institutes.<sup>12</sup> PET-based semi-automatically tumor contours were made by one observer using dedicated software (TrueD, Siemens Medical, Erlangen, Germany). Contours were defined by a threshold for the standardized uptake-value (SUV) based on the tumor-to-background signal ratio, with the gluteus muscle as reference background.<sup>15,16</sup> From the resulting tumor contour, maximal tumor diameter (MaxDiam), metabolic tumor volume (MTV), and maximal and mean SUV values within the MTV were calculated. SUV measures were corrected for blood glucose level.<sup>17</sup> The same variables were scored for the intermediate CRT PETCT scan and for each variable a response index (RI) was calculated. The RI is the relative difference between the value of the intermediate scan and pre-CRT scan and defined as  $(X_{\text{pre}} - X_{\text{intermediate}}) / X_{\text{pre}} \times 100\%$ .

### ***Pathological assessment***

Pathological complete response was defined as ypT<sub>0</sub>N<sub>0</sub>, extracted from the pathologic reports of surgical specimens. All other cases (ypT+ and/or ypN+) were considered non-responders. The specimens were not re-evaluated centrally but the pathology protocols were very similar between institutes (3-5 mm slices of rectum tumor, intensified evaluation on several blocks of tissue at the tumor site, evaluation on 2-3 sublevels when no tumor tissue was found in initial block). For the 11 wait-and-see patients, a pCR was assigned if the patient was locally recurrence free after 12 months of follow-up.

### ***Statistical analysis***

All statistical analyses were implemented and performed in MATLAB (version 7.1, MathWorks Inc., Natick, MA). Any missing values (Maastricht: 1.1%, Rome: 1.9%) in the datasets were substituted using the Expectation-Maximization method.<sup>18</sup> Datasets were pooled per institute on an individual patient level. Wilcoxon rank-sum tests were performed to test for association between a single variable and pCR. In the multivariate setting, logistic regression was applied to classify complete responders and non-responders, using the significant predictors from the univariate analysis as inputs. In the case of two very highly correlated input variables, only one was selected (using Spearman's correlation matrix,  $p < .05$ ). The model's accuracy was evaluated with the area under the curve (AUC) of the receiver operating characteristic (ROC) curve.<sup>19</sup> The maximum value of the AUC is 1.0, indicating a perfect prediction model while a value of 0.5 indicates a random chance of correct prediction. Predictors were only selected if their addition resulted in a sufficient AUC change ( $> .01$ ). For the final accuracy assessment, a bootstrapping scheme was applied (sampling with replacement,  $N=1000$ ), resulting in 95% confidence intervals for AUC. A nomogram was generated to represent a visualization of the final predictive model in which three risk groups were defined by applying two thresholds for the estimated probability for pCR. To define the low probability group for pCR, first the (weighted) average rate of non-responders (TRG3-4) was calculated from literature and thereafter a threshold was selected that resulted in this percentage of non-responders.<sup>12,20-24</sup> The threshold for the high estimated probability of pCR was calculated using decision curve analysis.<sup>25</sup> This method optimizes the threshold by calculating the net benefit of applying such a model and comparing it to the situations in which none or all patients are treated with a wait-and-see policy.

**Table 4.2** Patient characteristics of the training dataset (Maastricht) and the validation dataset (Rome).

		<b>Maastricht</b>	<b>Rome</b>
		<b>N</b>	<b>N</b>
		<b>[%]</b>	<b>[%]</b>
<b>Clinical</b>	<b>Sex</b>		
	Female	2 [25.9]	2 [35.9]
	Male	8 [74.1]	5 [64.1]
	<b>Age (years)</b>		
	Median	65.0	66.3
	Range	44.0 – 81.1	27.0 - 82.7
	<b>Clinical tumor stage</b>		
	2	1 [15.2]	5 [6.4]
	3	8 [76.8]	4 [62.8]
	4	9 [8.0]	2 [30.8]
<b>Clinical nodal stage</b>			
0	1 [13.4]	4 [5.1]	
+	9 [86.6]	7 [94.9]	
<b>PET imaging</b>	<b>Time between PET scans (days)</b>		
	Mean	21.9	28.5
	Standard deviation	± 2.5	± 10.5
	<b>Time 1<sup>st</sup> PET injection to acquisition (minutes)</b>		
	Mean	82.6	80.2
	Standard deviation	± 18.1	± 21.3
<b>Time 2<sup>nd</sup> PET injection to acquisition (minutes)</b>			
Mean	69.2	81.9	
Standard deviation	± 15.2	± 23.0	
<b>Treatment</b>	<b>Total radiotherapy dose (Gy)</b>		
	<50.4	5 [4.5]	2 [2.6]
	50.4	1 [95.5]	1 [19.2]
	55.0	0 [0.0]	6 [78.2]
	<b>Time last RT fraction to surgery (days)</b>		
Mean	73.6	72.9	
Standard deviation	± 18.8	± 13.2	
<b>Outcome</b>	<b>Pathologic complete response</b>		
	yes	2 [21.4]	1 [23.1]
	no	8 [78.6]	6 [76.9]

## Results

### *Dataset characterization*

Pooling the clinical and PETCT imaging data per institute resulted in similar cohort characteristics (Table 4.2). The validation dataset from Rome had in general: less males (64.1% vs 74.1%,  $p=.187$ ), higher clinical tumor stages ( $cT_4$ : 30.8% vs 8.0%,  $p<.001$ ), and more nodal involvement (94.9% vs 86.6%,  $p=.10$ ). Almost all Maastricht patients received 50.4Gy of RT (95.5%) while the majority of the Rome patients received 55.0Gy (78.2%). There was also a small but non-significant difference in the number of pathologic complete responders (23.1% vs 21.4%,  $p=.927$ ). The average time from last RT fraction to surgery was equal ( $73.6 \pm 18.8$  vs  $72.9 \pm 13.2$  days, mean  $\pm$  SD). Despite harmonization of the PET protocols in the THUNDER study, the time between the PET scans was on average lower for the Maastricht dataset ( $21.9 \pm 2.5$  vs  $28.5 \pm 10.5$  days). Also the time between tracer injection and PET acquisition showed was lower for the intermediate PET scan in Maastricht ( $69.5 \pm 15.2$  vs  $81.9 \pm 23.0$  minutes).

### *Predictor selection*

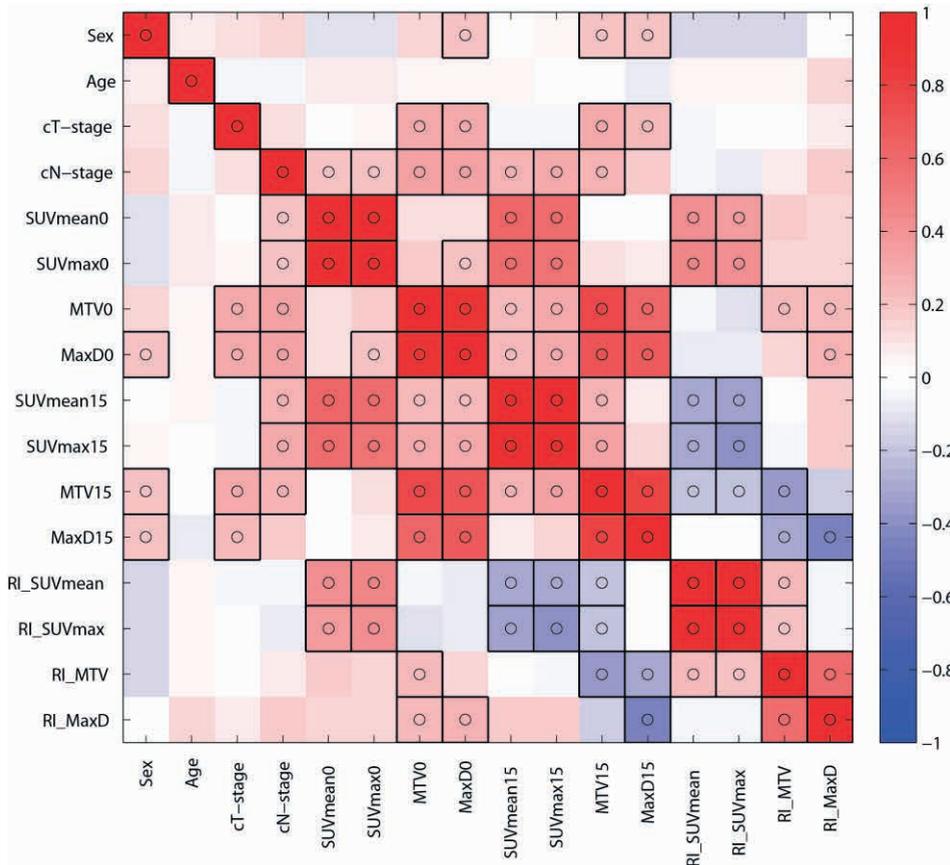
Univariate analyses (Table 4.3) showed that age, pre-treatment SUV measures, intermediate  $SUV_{mean}$ , and response index for maximal diameter have no significant predictive value for pCR ( $\alpha=.05$ ). Negatively correlated significant predictors (i.e. increasing value results in lower pCR rate) were cT-stage, cN-stage, pre-treatment and intermediate metabolic tumor volume (MTV) and maximal diameter, and intermediate  $SUV_{max}$ . Positively correlated significant predictors (i.e. increasing value results in higher pCR rate) were the response indices for  $SUV_{mean}$ ,  $SUV_{max}$  and MTV. Female gender was also found to be significantly associated with high pCR rate.

In multivariate logistic modeling only cT-stage was found significant in the total group of input predictors ( $p=.027^*$ ). However, highly correlated input variables increase p-values in a multivariate setting (Figure 4.2). The following decisions were made based on the analyses to select the final set of predictors:

- Gender was excluded: non-significant p-value and no other correlations with inputs
- cN-stage was included: significance near decision boundary ( $p=.056$ )
- The MTV measures were excluded: many outliers were detected (pre-treatment:  $N=11$  with MTV differences up to 4 times the average volume, intermediate:  $N=12$  with MTV differences up to 10 times the average volume). These measures also didn't have an added predictive value to the final selection.

- RI of  $SUV_{mean}$  was included at the expense of RI of  $SUV_{max}$ : literature reported sufficient evidence for both of them (Table 4.1). However, these measures are highly correlated (Figure 4.2). Univariate RI of  $SUV_{mean}$  had highest discriminative ability ( $p=.022^*$  vs  $p=.030^*$ ) and was therefore selected.
- From the other predictors, pre-treatment and intermediate maximal diameter and intermediate  $SUV_{max}$ , only intermediate maximal diameter was selected because it showed an AUC increase  $>0.02$  when added to the final set (and significance  $\alpha < 0.1$ ).

Hence, the final selected predictors in the multivariate model were cT-stage ( $p=.007^*$ ), cN-stage ( $p=.032^*$ ), intermediate maximal diameter ( $p=.078$ ) and RI of  $SUV_{mean}$  ( $p=.025^*$ ).



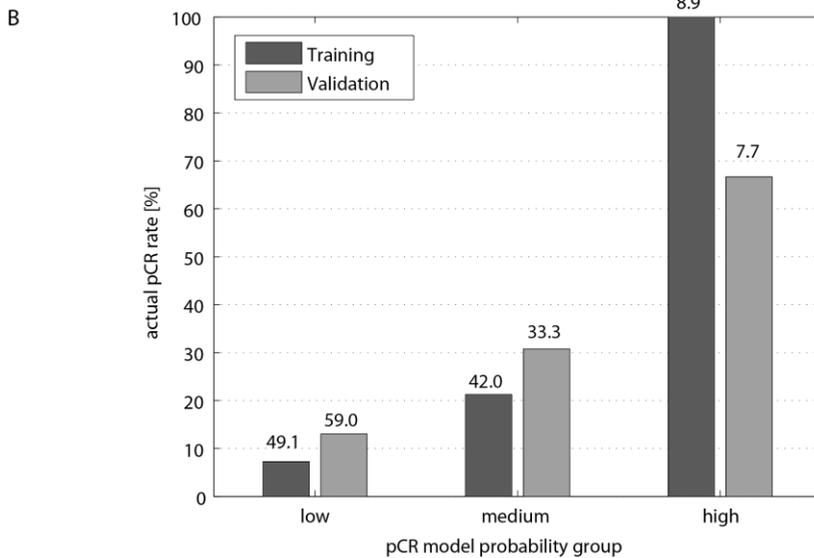
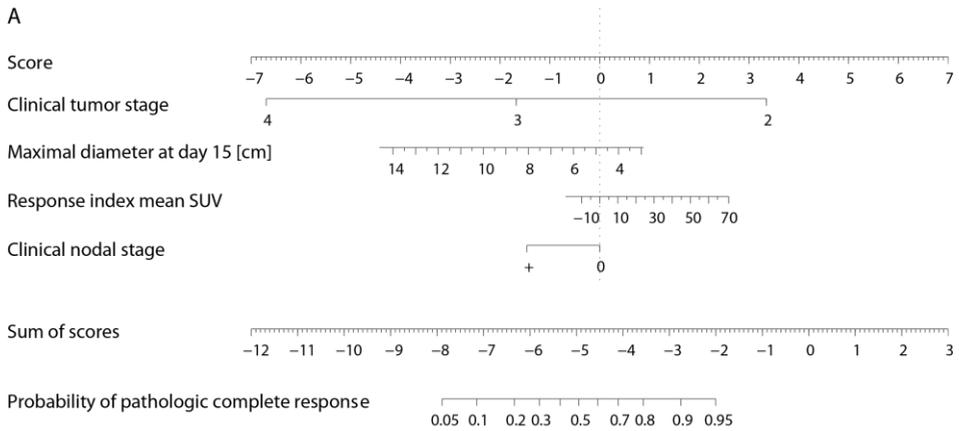
**Figure 4.2** Spearman correlation matrix to identify significant (boxed + inner circle) correlations between model input variables. Red indicates positive correlation, blue indicates negative correlation.

**Table 4.3** Prediction results. For each variable, the distributions are compared between training and validation dataset and a univariate analysis is performed (training set only). Multivariate analysis including feature selection are included.

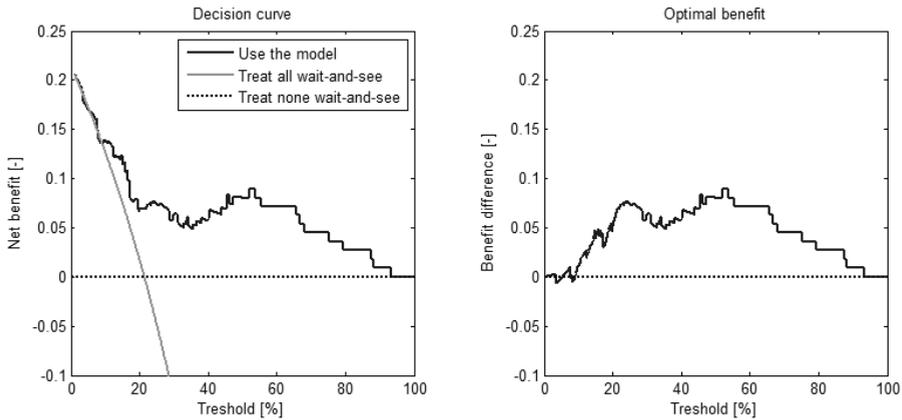
	Univariate		Multivariate			Predictor selection		
	p	pCR ↑	OR	[95% CI]	p	OR	[95% CI]	p
Gender	.048*	Female	0.65	[0.19-2.23]	.495			
Age	.523	-						
cT-stage	.002*	↓	0.20	[0.05-0.83]	.027*	0.19	[0.06-0.64]	.007*
cN-stage	.001*	↓	0.21	[0.04-1.04]	.056	0.23	[0.06-0.88]	.032*
SUV <sub>mean0</sub>	.747	-						
SUV <sub>max0</sub>	.617	-						
MTV <sub>0</sub>	.002*	↓	1.08	[1.00-1.17]	.061			
MaxDiam <sub>0</sub>	.004*	↓	0.90	[0.42-1.90]	.778			
SUV <sub>mean15</sub>	.067	-						
SUV <sub>max15</sub>	.030*	↓	0.96	[0.83-1.10]	.545			
MTV <sub>15</sub>	<.001*	↓	0.82	[0.67-1.02]	.072			
MaxDiam <sub>15</sub>	.005*	↓	0.95	[0.58-1.56]	.849	0.74	[0.53-1.03]	.078
RI_SUV <sub>mean</sub>	.022*	↑	1.01	[0.90-1.13]	.893	1.04	[1.00-1.07]	.025*
RI_SUV <sub>max</sub>	.030*	↑	1.01	[0.91-1.13]	.839			
RI_MTV	.017*	↑	0.99	[0.96-1.03]	.757			
RI_MaxDiam	.544	-						

### Validation of the nomogram

The multivariate model with the four selected predictors to estimate the probability of a pCR was visually represented by a nomogram (Figure 4.3A). Bootstrapped AUCs were 0.78 (95% CI: 0.65-0.89) for the training dataset and 0.70 (95% CI: 0.55-0.84) for the validation dataset. With the aim of estimating three probability groups for pCR, two probability thresholds were defined to separate these groups. The 12.8% threshold (<12.8% low probability on pCR, >12.8% medium probability of pCR) was defined based on literature in which on average 49.2% of the patients are non-responders (weighted for number of patients). The other threshold was calculated at 53% based on decision curve analysis where the optimal net benefit of applying a wait-and-see policy was maximized (Figure 4.4). These three probability groups (low, medium, high) resulted in significantly increasing pCR rates (training data: respectively 7.3%, 21.3%, 100% pCR; validation data: 13.0%, 30.8% and 66.7% pCR). The highest probability groups contained 8.9% (training) and 7.7% (validation) of the total number of patients, while the lowest probability groups contained respectively 49.1% and 59.0% (Figure 4.3B).



**Figure 4.3** (A) Nomogram of pathologic complete response based on the multivariate analysis. Three probability groups are defined and pCR rates of those groups are plotted. (B) Actual pCR rates in the training and validation set plotted for three probability groups from the model, which are defined by low:  $P \leq 12.8\%$ , medium:  $12.8\% < P < 53\%$ , high:  $P \geq 53\%$ . Relative number of patients in the group (%) is shown above the bars.



**Figure 4.4** Clinical usefulness assessment for the model to determine the optimal threshold to define the high probability group. This group is a candidate for the wait-and-see policy. Net benefit of the model is compared to the situations in which all patients or no patients are treated with a wait-and-see policy (left). The threshold with the optimal benefit of the model is 53% (right).

## Discussion

In this study a multivariate nomogram with clinical parameters and early sequential PETCT imaging markers predicting pCR in LARC was developed based on a large prospective study and validated prospectively within another institute. Good performances were measured for both training and validation dataset. After risk group identification a subgroup of just below 10% of the patients with high estimated probability for complete response was identified.

### *Model predictors*

The selected predictive factors for pCR were cT-stage, cN-stage,  $\Delta\text{SUV}_{\text{mean}}$  and intermediate maximal tumor diameter. A recent large analysis of 3105 patients found that cT-stage was predictive for pCR ( $p < .0001$ ) but that cN-stage shows only a trend ( $p = .10$ ).<sup>4</sup> This study contained however also old cases where CT was used for cN-stage scoring. Another analysis of 677 patients associated both cT-stage ( $p < .001$ ) and cN-stage ( $p < .001$ ) with pCR.<sup>10</sup> The same study also found that pre-treatment (metabolic) tumor size was predictive ( $p = .003$ ), but others show that only changes in metabolic volume in the pre-post treatment setting were significant for pCR and not the intermediate case ( $p = .010$ ).<sup>26</sup> In the presented study's univariate analysis, it is clear that tumor size is important for pCR. The change in  $\text{SUV}_{\text{mean}}$  at the intermediate time

point was found predictive in earlier studies for responders versus non-responders,<sup>12,20,21</sup> but  $SUV_{max}$  is stronger when predictions for pCR changes are made.<sup>22,27</sup> In our study the response index of  $SUV_{mean}$  was a stronger predictor than the RI of  $SUV_{max}$ . These two measures are also highly correlated, especially since our PET contouring was semi-automatic and calibrated for both institutes, resulting in less variation in  $SUV_{mean}$  due to contouring.<sup>28</sup>

A variable selection scheme was chosen based on univariate analysis, correlation between input variables and contribution to multivariate prediction performance. Other strategies such as penalized feature selection are also common when dealing with highly correlated variables, but in general these results in interpretability issues for the (shrunk) coefficients.<sup>29</sup>

### ***Model performance***

The performance of the nomogram measured by AUC of 0.78 (95% CI: 0.65-0.89) for the training dataset and 0.70 (95% CI: 0.55-0.84) for the validation dataset are lower than the ones reported by single PET parameters in literature (0.70,<sup>30</sup> 0.83,<sup>23</sup> 0.87<sup>12</sup>). However, these studies predict response (TRG1-2) instead of pCR. Response prediction is in practice more accurate because the number of events for good response (45-55%) is much higher than those of pCR (15-30%). Identifying the complete responders in an early phase is useful to avoid additional treatments and related toxicity for these patients. Another possible reason for a lower overall performance can be the noisy pathology outcome (decentralized), but this is compensated by the high number of patients. The current studies reported in literature with low number of samples are sensitive to positive (or negative) findings by mere chance, and therefore it is reasoned that this large study's performance is expected to reflect reality better.

When stratifying the patients in three risk groups, the performance is acceptable: 100% accuracy for high probability for pCR group for the training dataset (N=10) and 67% for the validation dataset (N=6). The two misclassified patients in the validation set have TRG 2, ypT<sub>1</sub>N<sub>0</sub> or ypT<sub>2</sub>N<sub>0</sub> status and a clinical response on the PETCT scan two months after the end of CRT, thus they are considered as good responders.

The difference in performance between training and validation dataset is likely to be based on differently distributed data with respect to the predictors. The validation set has lower model estimates for the probabilities of pCR due to significantly higher number pre-treatment cT-stages (p=.007), and significantly lower RIs of  $SUV_{mean}$  (p<.001). The latter difference may be explained by the higher times between tracer injection and the intermediate PET acquisition in the validation institute in comparison to the training institute, resulting in higher intermediate  $SUV_{mean}$  values (p<.001), despite PET protocol harmonization.

### ***Model applicability***

As suggested, the developed model can be used to assist in decision making for LARC already during CRT (Figure 4.1). However, three notes have to be made. First, this model is only useful for decisions made during or immediately after CRT, like RT boost or additional chemotherapy. The decision for a wait-and-see approach can better be made just before surgery by using both specialized prediction models<sup>10</sup> and careful assessment of imaging, endoscopies and biopsies.<sup>5</sup> The advantages of an earlier estimate of pCR are avoidance of overtreatment of complete responders, a possible increase of the number of complete responders with a RT boost for good responders and perhaps a change in treatment strategy for non-responding patients.

Secondly, any developed model requires prospective validation by means of a randomized trial, comparing an arm with standard treatment for all (CRT+surgery) to an arm receiving individualized treatment based on the prediction model. Such a study is currently being set up.

And last, other predictors from different sources might be considered to further improve accuracy. Diffusion-weighted magnetic resonance imaging (DW-MRI) at different time points is reported as a promising candidate which increases the prediction accuracy significantly in combination with PETCT imaging.<sup>27</sup> Blood biomarkers also can have additional value as for example has been reported for serum carcinoembryonic antigen (CEA).<sup>31</sup> For all these additional sources however, cost-benefit analyses are advised because saturation of the prediction accuracy can become an issue.

### ***Conclusions***

We have developed an externally validated and accurate prediction model for pathologic complete response in locally advanced rectal cancer based on large prospective studies. This nomogram can be used to distinguish three types of patients, i.e. complete responders, good responders and non-responders, for which respectively a wait-and-see policy, radiotherapy boost and additional chemotherapy can be administered. This personalized treatment approach is expected to promote more complete responders, reduce the number of surgeries and related complications, and to avoid unnecessary toxicity.

***Acknowledgements***

Authors acknowledge financial support from the QuIC-ConCePT project, which is partly funded by EFPI A companies and the Innovative Medicine Initiative Joint Undertaking (IMI JU) under Grant Agreement No. 115151. Authors also acknowledge financial support from the National Institute of Health (NIH-USA U01 CA 143062-01), the CTMM framework (AIRFORCE project, grant 030-103), EU 7th framework program (METOXIA, EURECA), euroCAT (IVA Interreg - [www.eurocat.info](http://www.eurocat.info)) and the Dutch Cancer Society (KWF UM 2011-5020, KWF UM 2009-4454).

## References

1. Lambin P, van Stiphout RG, Starmans MH, et al. Predicting outcomes in radiation oncology--multifactorial decision support systems. *Nat Rev Clin Oncol* 2013; 10:27-40
2. Valentini V, Lambin P, Myerson RJ. Is it time for tailored treatment of rectal cancer? From prescribing by consensus to prescribing by numbers. *Radiother Oncol* 2012; 102:1-3
3. Gerard JP, Conroy T, Bonnetain F, et al. Preoperative radiotherapy with or without concurrent fluorouracil and leucovorin in T3-4 rectal cancers: results of FFCD 9203. *J Clin Oncol* 2006; 24:4620-5
4. Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 2010; 11:835-44
5. Maas M, Beets-Tan RG, Lambregts DM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011; 29:4633-40
6. Gerard JP, Chapet O, Nemoz C, et al. Improved sphincter preservation in low rectal cancer with high-dose preoperative radiotherapy: the lyon R96-02 randomized trial. *J Clin Oncol* 2004; 22:2404-9
7. Braendengen M, Tveit KM, Berglund A, et al. Randomized phase III study comparing preoperative radiotherapy with chemoradiotherapy in nonresectable rectal cancer. *J Clin Oncol* 2008; 26:3687-94
8. Zhang C, Tong J, Sun X, et al. 18F-FDG-PET evaluation of treatment response to neo-adjuvant therapy in patients with locally advanced rectal cancer: a meta-analysis. *Int J Cancer* 2012; 131:2604-11
9. Strauss LG. Positron Emission Tomography: Current Role for Diagnosis and Therapy Monitoring in Oncology. *Oncologist* 1997; 2:381-388
10. van Stiphout RG, Lammering G, Buijsen J, et al. Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential PET-CT imaging. *Radiother Oncol* 2011; 98:126-33
11. Janssen MH, Ollers MC, Riedl RG, et al. Accurate prediction of pathological rectal tumor response after 2 weeks of pre-operative radiochemotherapy using FDG-PET-CT imaging. *Int J Radiat Oncol Biol Phys* 2009;
12. Janssen MH, Ollers MC, van Stiphout RG, et al. PET-based treatment response evaluation in rectal cancer: prediction and validation. *Int J Radiat Oncol Biol Phys* 2012; 82:871-6
13. Barbaro B, Vitale R, Valentini V, et al. Diffusion-weighted magnetic resonance imaging in monitoring rectal cancer response to neoadjuvant chemoradiotherapy. *Int J Radiat Oncol Biol Phys* 2012; 83:594-9
14. Engelen SM, Beets-Tan RG, Lahaye MJ, et al. Location of involved mesorectal and extramesorectal lymph nodes in patients with primary rectal cancer: preoperative assessment with MR imaging. *Eur J Surg Oncol* 2008; 34:776-81
15. Daisne JF, Sibomana M, Bol A, et al. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. *Radiother Oncol* 2003; 69:247-50
16. Ollers M, Bosmans G, van Baardwijk A, et al. The integration of PET-CT scans from different hospitals into radiotherapy treatment planning. *Radiother Oncol* 2008; 87:142-6
17. Janssen MH, Ollers MC, van Stiphout RG, et al. Blood glucose level normalization and accurate timing improves the accuracy of PET-based treatment response predictions in rectal cancer. *Radiother Oncol* 2010; 95:203-8
18. Ludbrook J. Outlying observations and missing values: how should they be handled? *Clin Exp Pharmacol Physiol* 2008; 35:670-8

19. Pepe MS. Receiver Operating Characteristic Methodology. *Journal of the American Statistical Association* 2000; 95
20. Avallone A, Aloj L, Caraco C, et al. Early FDG PET response assessment of preoperative radiochemotherapy in locally advanced rectal cancer: correlation with long-term outcome. *Eur J Nucl Med Mol Imaging* 2012; 39:1848-57
21. Cascini GL, Avallone A, Delrio P, et al. 18F-FDG PET is an early predictor of pathologic tumor response to preoperative radiochemotherapy in locally advanced rectal cancer. *J Nucl Med* 2006; 47:1241-8
22. Goldberg N, Kundel Y, Purim O, et al. Early prediction of histopathological response of rectal tumors after one week of preoperative radiochemotherapy using 18 F-FDG PET-CT imaging. A prospective clinical study. *Radiat Oncol* 2012; 7:124
23. Guerra L, Niespolo R, Di Pisa G, et al. Change in glucose metabolism measured by 18F-FDG PET/CT as a predictor of histopathologic response to neoadjuvant treatment in rectal cancer. *Abdom Imaging* 2011; 36:38-45
24. Leibold T, Akhurst TJ, Chessin DB, et al. Evaluation of (1)(8)F-FDG-PET for early detection of suboptimal response of rectal cancer to preoperative chemoradiotherapy: a prospective analysis. *Ann Surg Oncol* 2011; 18:2783-9
25. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26:565-74
26. Sun W, Xu J, Hu W, et al. The role of sequential 18(F) -FDG PET/CT in predicting tumour response after preoperative chemoradiation for rectal cancer. *Colorectal Dis* 2013; 15:e231-8
27. Lambrecht M, Deroose C, Roels S, et al. The use of FDG-PET/CT and diffusion-weighted magnetic resonance imaging for response prediction before, during and after preoperative chemoradiotherapy for rectal cancer. *Acta Oncol* 2010; 49:956-63
28. Buijssen J, van den Bogaard J, van der Weide H, et al. FDG-PET-CT reduces the interobserver variability in rectal tumor delineation. *Radiother Oncol* 2012; 102:371-6
29. Ma S, Huang J. Penalized feature selection and classification in bioinformatics. *Brief Bioinform* 2008; 9:392-403
30. Rosenberg R, Herrmann K, Gertler R, et al. The predictive value of metabolic response to preoperative radiochemotherapy in locally advanced rectal cancer measured by PET/CT. *Int J Colorectal Dis* 2009; 24:191-200
31. Moureau-Zabotto L, Farnault B, de Chaisemartin C, et al. Predictive factors of tumor response after neoadjuvant chemoradiation for locally advanced rectal cancer. *Int J Radiat Oncol Biol Phys* 2011; 80:483-91

# Chapter 5

Blood biomarkers are helpful in the prediction of response to chemoradiation in rectal cancer: a prospective, hypothesis driven study on patients with locally advanced rectal cancer

Ruud van Stiphout\*, Jeroen Buijsen\*, Paul Menheere, Guido Lammering, Philippe Lambin

Submitted

## Abstract

### *Purpose*

Chemoradiation (CRT) has been shown to lead to downsizing in an important part of rectal cancers. In order to tailor treatment at an earlier stage during treatment, predictive models are being developed. Adding blood biomarkers may be attractive for prediction, as they can be collected very easily and determined with excellent reproducibility in clinical practice. The hypothesis of this study was that blood biomarkers related to tumor load, hypoxia and inflammation can help to predict response to CRT in rectal cancer.

### *Methods and materials*

295 patients with locally advanced rectal cancer who were planned to undergo CRT were prospectively entered into a biobank protocol (NCT01067872). Blood samples were drawn before start of CRT. Nine biomarkers were selected, based on a previously defined hypothesis, and measured in a standardized way by a certified lab: CEA, CA19-9, LDH, CRP, IL-6, IL-8, CA IX, osteopontin and 25-OH-vit D. Outcome was analyzed in two ways: pCR vs. non-pCR and responders (defined as ypT<sub>0-2</sub>N<sub>0</sub>) vs. non-responders (all other ypTN stages).

### *Results*

276 patients could be analyzed. 20.7% developed a pCR and 47.1% were classified as responder. In univariate analysis CEA ( $p=.001$ ) and osteopontin ( $p=.012$ ) were significant predictors for pCR. Taking response as outcome CEA ( $p<.001$ ), IL-8 ( $p<.001$ ) and osteopontin ( $p=.004$ ) were significant predictors. In multivariate analysis CEA was the strongest predictor for pCR (OR 0.92,  $p=.019$ ) and CEA and IL-8 predicted for response (OR 0.97,  $p=.029$  and OR 0.94,  $p=.036$ ). The model based on biomarkers only had an AUC of 0.65 for pCR and 0.68 for response; the strongest model included clinical data, PET-data and biomarkers and had an AUC of 0.81 for pCR and 0.78 for response.

### *Conclusions*

CEA and IL-8 were identified as predictive biomarkers for tumor response and pCR after CRT in rectal cancer. Incorporation of these blood biomarkers lead to an additional accuracy of earlier developed prediction models using clinical variables and PET-information. The new model could help to an early adaptation of treatment in rectal cancer patients.

## Introduction

Combined treatment is the cornerstone of rectal cancer treatment. In case of locally advanced rectal cancer, defined as a tumor with a predicted positive circumferential resection margin (CRM) or four or more positive lymph nodes, chemoradiotherapy (CRT) has become standard of care.<sup>1</sup> Pathological complete response (pCR) rates typically lie between 15 and 20% after CRT depending on the radiotherapy dose given and the interval between CRT and surgery.<sup>2,3</sup> The group of patients that develop a pCR is of particular interest, because they have a better prognosis<sup>2</sup> and may be offered less invasive surgery<sup>4</sup> or surgery may be completely omitted<sup>5,6</sup>. Therefore it would be an advantage if the pCR rate could be increased. There are several possible strategies, of which early response prediction during CRT, leading to treatment adaptation, is very attractive.

In the past, clinical parameters as well as information from PET-scans before and during treatment have been found to be predictive for treatment outcome.<sup>7-9</sup> A prediction model based on tumor length, cT- and cN-stage had a predictive performance of 0.61 as expressed by the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. A second model including maximal standardized uptake value (SUV) of the tumor derived from a PET-scan before the start of treatment, maximal tumor diameter as measured on PET-scan, tumor location and cN-stage, resulted in an AUC of 0.68.<sup>9</sup> PET-scan after 2 weeks of CRT has been shown to be very predictive for response (tumor regression grade (TRG) 1-2 vs TRG 3-5 according to Mandard).<sup>8</sup>

It is attractive to consider the addition of blood biomarkers to these predictors, since samples can be collected easily, are relatively cheap to measure, and they contain information about different aspects of tumor biology. Furthermore, they can be measured accurately and precisely using standardized methods. Reports on the predictive value of blood biomarkers are limited to studies mainly evaluating 1 or 2 biomarkers.<sup>7,10-13</sup> The most studied biomarker for response to CRT in rectal cancer is carcinoembryonic antigen (CEA) and one study analyzed the predictive value of osteopontin and interleukin-6. Based on these data combined with data of prognostic studies in colorectal cancer and our experience with a blood biomarker model in lung cancer,<sup>14</sup> we decided to include 9 biomarkers. CEA and CA19-9 are related to tumor load, interleukin-6 and -8 (IL-6 and IL-8) and C-reactive protein (CRP) are markers of inflammation, lactate dehydrogenase (LDH) is a marker of cell death, carbonic anhydrase IX (CA IX) and osteopontin are hypoxia markers and 25-OH-vitamin D may induce growth arrest and apoptosis of tumor cells.

In this prospective study we tested the hypothesis that these biomarkers are helpful in the prediction of response to CRT in rectal cancer.

## Methods and materials

### *Patient population*

We identified 295 patients who were treated with CRT with curative intent for locally advanced rectal cancer between January 2005 and December 2009. All gave written informed consent to be included in a biobank protocol (NCT01067872). Nineteen patients were ineligible for different reasons (1 patient was treated with short course radiotherapy 5x5 Gy, 5 patients were not operated on and entered in a wait&see study so no ypTN stage could be determined, 2 patients underwent a TEM resection so no ypN-stage was available, 6 patients had metastases and were treated with palliative intent, 1 patient died during treatment, in 1 patient all biobank material was hemolytic and in 3 patients the PA report could not be retrieved), resulting in 276 patients for analysis. In 9 patients there were technical problems with the biomarker measurements, so that 267 patients were available for the biomarker analysis. Locally advanced disease was defined as a distal T<sub>3</sub> tumor and/or N<sub>2</sub> status and/or a mid- or upper-rectal tumor with a predicted circumferential resection margin <2 mm, or any T<sub>4</sub> tumor. Locoregional staging for clinical tumor and nodal stage was based on MRI. In 198 patients for whom biomarkers were available a PETCT was made for radiotherapy planning. Treatment consisted of 28 fractions of 1.8 Gy on the primary tumor, mesorectum, presacral area and external iliac lymph nodes in combination with capecitabine 825 mg/m<sup>2</sup> BID. Patients were operated on 8-10 weeks after the end of CRT. Pathology reports were collected from the referring hospitals.

### *Blood samples*

Blood samples were collected before the start of treatment. Samples were processed and stored using a standard protocol. All biomarkers were measured in serum, except for osteopontin and CA IX, which were measured in EDTA plasma. Biomarker measurements were done in one single, certified laboratory, using commercially available kits. All samples were analyzed simultaneously. Measurements were performed using the following kits: CEA was measured using a solid-phase, two-site sequential chemoluminescent immunometric assay (Siemens Medical Solutions Diagnostics, LA, USA), IL-6 and IL-8 were determined with a solid phase, enzyme labeled, chemoluminescence sequential immunometric assay (Siemens Medical

Solutions Diagnostics, LA, USA). LDH (Beckman Coulter, Fullerton, CA), CRP (Beckman, Coulter Fullerton, CA) CA 19-9 has been determined on Brahms Kryptor (Brahms, ThermoFisher, Hennigsdorf, Germany) with a sandwich immuno-fluorescent assay. 25-OH-Vitamine-D was measured with a commercially available radioimmunoassay (IDS, Frankfurt am Main, Germany). CA IX is measured by an enzyme-linked immunosorbent assay (Wilex (OncogenScience), Cambridge, MA, USA), and OPN was measured by an enzyme-linked immunosorbent assay (Quantikine Human Osteopontin Immuno assay; R&D Systems, Minneapolis, MN, USA). OPN, 25-OH-vitamin-D and CA IX were measured using manual methods in duplicate. All other biomarkers were measured in singletons.

### ***Statistical analysis***

Outcome after CRT was analyzed in 2 different ways: 1) pCR, defined as the absence of any tumor cells in the operative pathologic specimen, at the primary site, or in lymph node regions, versus non-pCR and 2) good responders, defined as ypT<sub>0-2</sub>N<sub>0</sub>, versus poor responders, defined as ypT<sub>3</sub> and/or ypN<sub>1-2</sub>. Missing values were completed using expectation-maximization imputation. Correlations between biomarkers were analyzed by calculating Spearman's rho. Because the biomarkers showed a skewed distribution, the Mann-Whitney *U* test was used to determine significant univariate predictors of response. Logistic regression analysis was used to identify the significant multivariate predictors of response. The next step was to combine blood biomarkers with clinical data and PET parameters. For this analysis clinical and PET-variables were selected manually, based on an earlier predictive model for pCR.<sup>9</sup> The included clinical variables were tumor length, clinical T and N stage, all based on MRI. Included PET-features were maximum SUV, pre-treatment metabolic volume and maximum diameter. The two latter variables were measured using a source-to-background ratio method as has been described earlier.<sup>15</sup> ROC curves were constructed and the AUC was calculated. In order to approximate the true AUC and calculate confidence intervals bootstrapping (n=1000) was used. A perfect prediction model results in an AUC of 1.0, while an AUC of 0.5 indicates that response is predicted correctly in 50% of cases (*i.e.* as good as chance). Statistical analyses were performed using Matlab, release 2010b (The MathWorks, Natick, MA).

### ***Ethics***

The biobank study was conducted according to the Dutch law and approved by the local medical ethics committee. All patients gave written informed consent before collection of the blood samples.

## Results

The patient characteristics are shown in Table 5.1. The majority of patients had a tumor penetrating through the bowel wall and predicted positive lymph nodes on MR. In the total database 20.7% of patients developed a pCR and 47.1% of patients were classified as responder. Table 5.2 shows the results of the biomarker measurements for the different outcome groups, as well as the PET-parameters that were included in the model. In general lower serum levels for blood biomarkers were seen in the poor responding groups (except for 25-OH-vitamin-D). Additional analysis (not shown in the table) revealed significant positive correlations between IL-6, IL-8, CRP and osteopontin and between CEA and CA19-9.

**Table 5.1** Patient characteristics (N = 276)

Characteristic	N	[%]
<b>Age [years]</b>		
median	65.8	
range	23.0 - 92.2	
<b>Gender</b>		
male	179	[64.9]
female	97	[35.1]
<b>Clinical tumor stage (cT)</b>		
2	26	[9.4]
3	207	[75.0]
4	43	[15.6]
<b>Clinical nodal stage (cN)</b>		
0	47	[17.0]
1	114	[41.3]
2	115	[41.7]
<b>WHO performance index</b>		
0	218	[79.0]
1	53	[19.2]
2	5	[1.8]
<b>Tumor length [cm]</b>		
median	5.0	
range	2.0 - 13.0	
<b>pCR (ypT<sub>0</sub>N<sub>0</sub>)</b>		
no	219	[79.3]
yes	57	[20.7]
<b>Good response (ypT<sub>012</sub>N<sub>0</sub>)</b>		
no	146	[52.9]
yes	130	[47.1]

### Univariate analysis

Univariate analysis indicated that CEA and osteopontin were significant predictors for pCR ( $p=.001$ ,  $p=.012$  respectively) and that CEA, IL-8 and osteopontin were significant predictors for response ( $p<.001$ ,  $p<.001$ ,  $p=.004$  respectively) as shown in Table 5.3. Lower serum levels of these markers correlated with a higher chance of response to chemoradiation. Of the clinical parameters clinical N-stage was predictive for pCR ( $p=.026$ ) and response ( $p=.001$ ) in univariate analysis and tumor length ( $p=.02$ ) and clinical T-stage ( $p=.004$ ) for response only. The pre-treatment metabolic volume and maximum diameter based on PET were predictive for both outcome measures ( $p=.016$  and  $0.009$  for pCR and  $p=.006$  and  $.005$  for response respectively).

**Table 5.2** Biomarker levels and PET parameters (average  $\pm$  standard deviation) compared for the subpopulations of pCR vs no pCR and good response vs no good response

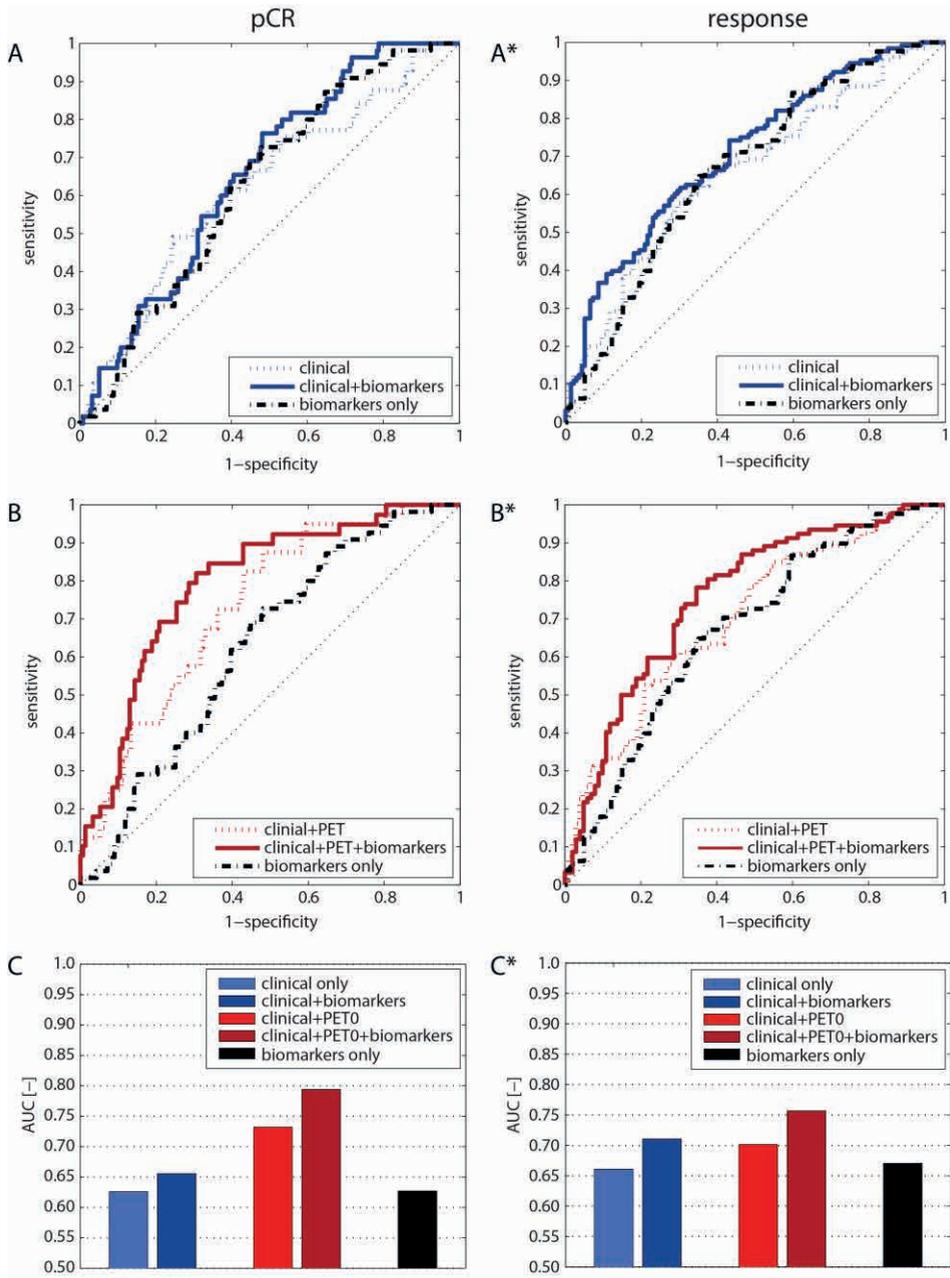
		pCR (ypT <sub>0</sub> N <sub>0</sub> )		Good response (ypT <sub>0/12</sub> N <sub>0</sub> )	
		no	yes	no	yes
Biomarkers (N = 267)	CEA	14.7 $\pm$ 28.4	8.7 $\pm$ 22.9	18.1 $\pm$ 33	8.5 $\pm$ 18.5
	IL-6	4.0 $\pm$ 8.4	2.9 $\pm$ 2.7	4.8 $\pm$ 10.1	2.7 $\pm$ 2.4
	IL-8	15.8 $\pm$ 9	14.0 $\pm$ 7.4	17.3 $\pm$ 10	13.4 $\pm$ 6.4
	LDH	181.3 $\pm$ 43.8	176.2 $\pm$ 33.9	181.9 $\pm$ 46.7	178.5 $\pm$ 36.1
	CRP	11.5 $\pm$ 24.5	7.8 $\pm$ 8	13.8 $\pm$ 29.1	7.5 $\pm$ 9.2
	CA 19-9	26.9 $\pm$ 29.1	25.1 $\pm$ 31.9	28.3 $\pm$ 28.8	24.6 $\pm$ 30.6
	vitD-25	53.2 $\pm$ 20.9	55.5 $\pm$ 18.8	52.6 $\pm$ 22.4	54.8 $\pm$ 18.2
	CA-9	282.5 $\pm$ 275	274.4 $\pm$ 302.9	282.0 $\pm$ 238.4	279.6 $\pm$ 320.9
	OPN	79.2 $\pm$ 28.9	68.2 $\pm$ 16	81.6 $\pm$ 30.9	72 $\pm$ 21.3
PET (N =198)	SUV <sub>max</sub>	15.3 $\pm$ 6.3	13.6 $\pm$ 5.2	14.3 $\pm$ 5.6	15.6 $\pm$ 6.7
	SUV <sub>mean</sub>	8.2 $\pm$ 3.1	7.4 $\pm$ 2.7	7.8 $\pm$ 2.7	8.4 $\pm$ 3.3
	MTV [cc]	33.8 $\pm$ 33.5	25 $\pm$ 20.2	37.6 $\pm$ 39.1	25.7 $\pm$ 17.6
	Max diam [cm]	6.5 $\pm$ 1.9	5.8 $\pm$ 1.6	6.7 $\pm$ 2.1	5.9 $\pm$ 1.4

### Multivariate biomarker model

Table 5.3 shows the results of the multivariate analysis for the total set of parameters as well as a selection of biomarkers and clinical and PET-parameters. In the complete set of parameters IL-8 was the only significant predictor for response ( $p=.05$ ), while osteopontin was borderline significant for pCR prediction ( $p=.056$ ). As a next step a manual selection of the most promising predictors was made. Blood biomarkers that had a significant predictive value in either univariate or multivariate analysis were included and IL-6, although not significant, was included because it had a predictive value in a prognostic model for lung cancer.<sup>14</sup> In this selection of biomarkers consisting

Table 5.3 Statistical analysis to show association of the candidate variables with pCR and response

	Univariate		Multivariate total				Multivariate selection			
	pCR	Response	pCR	Response	pCR	Response	pCR	Response	pCR	Response
	P	P	OR [95% CI]	P	OR [95% CI]	P	OR [95% CI]	P	OR [95% CI]	P
Age	.282	.775	1.0 [0.96-1.03]	.956	1.01 [0.98-1.04]	.395				
Sex	.224	.916	1.65 [0.84-3.26]	.147	1.18 [0.66-2.11]	.57				
WHO	.353	.168	0.95 [0.44-2.06]	.894	0.97 [0.52-1.78]	.913				
Tumor length	.404	.020*	1.01 [0.85-1.19]	.925	0.94 [0.82-1.08]	.388			1.0 [0.79-1.27]	.987
cT-stage	.096	.004*	0.69 [0.35-1.39]	.299	0.53 [0.29-0.96]	.035*			0.28 [0.1-0.78]	.015*
cN-stage	.026*	.001*	0.68 [0.44-1.04]	.076	0.58 [0.4-0.84]	.005*			0.43 [0.24-0.78]	.006*
CEA	.001*	<.001*	0.99 [0.97-1.01]	.323	0.99 [0.97-1.0]	.095			0.92 [0.85-0.99]	.019*
IL-6	.487	.119	1.0 [0.91-1.09]	.928	0.95 [0.87-1.03]	.196			0.99 [0.89-1.1]	.878
IL-8	.107	<.001*	1.0 [0.96-1.05]	.991	0.96 [0.92-1.0]	.05*			1.01 [0.96-1.07]	.746
LDH	.637	.840	0.99 [0.98-1]	.162	1.0 [0.99-1.0]	.21				
CRP	.875	.291	1.0 [0.98-1.02]	.937	1.0 [0.98-1.02]	.75				
CA 19-9	.245	.103	1.0 [0.99-1.01]	.723	1.0 [0.99-1.01]	.998				
vitD-25	.348	.147	1.01 [0.99-1.02]	.523	1.0 [0.99-1.02]	.547				
CA-9	.564	.134	1.0 [1.0-1.0]	.570	1.0 [1.0-1.0]	.706				
OPN	.012*	.004*	0.98 [0.96-1]	.056	1.0 [0.98-1.01]	.433			0.98 [0.96-1.01]	.136
SUV <sub>max0</sub>	.335	.193	0.92 [0.64-1.33]	.665	1.02 [0.78-1.32]	.91			0.94 [0.87-1.02]	.141
SUV <sub>mean0</sub>	.316	.153	1.08 [0.52-2.21]	.844	1.09 [0.64-1.87]	.74				
GTV <sub>0</sub>	.016*	.006*	1.0 [0.97-1.03]	.965	0.98 [0.96-1.01]	.226				
MaxD <sub>0</sub>	.009*	.005*	0.99 [0.58-1.7]	.967	0.98 [0.65-1.5]	.94				

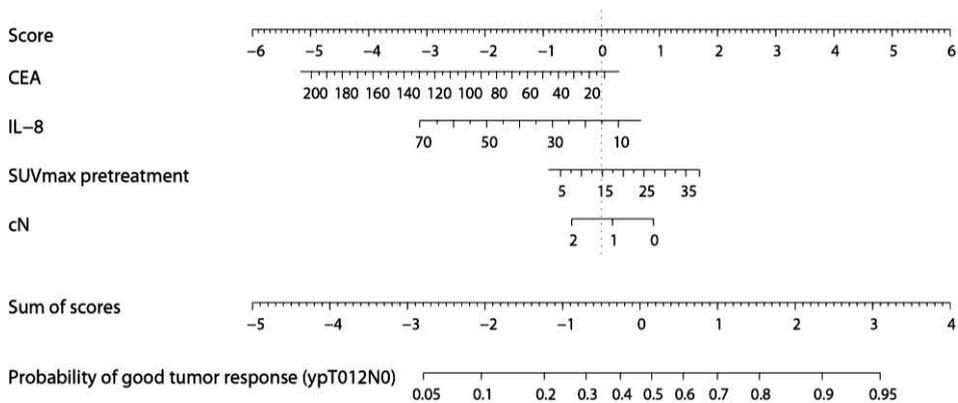


**Figure 5.1** ROC curves for prediction models for pCR and response including biomarkers and clinical variables (A and A\*) and biomarkers, clinical variables and PET-data (B and B\*) and resulting AUC for the different prediction models (C and C\*).

of CEA, IL-6, IL-8 and osteopontin, CEA was the only significant predictor of pCR in multivariate analysis ( $p=.019$ ) and CEA and IL-8 significantly predicted response ( $p=.029$ ,  $p=.021$  respectively). Including all biomarkers resulted in an AUC of 0.65 (95% CI 0.57-0.73) for pCR prediction and 0.68 (95% 0.61-0.75) for response prediction. Table 5.3 shows the odds ratios for all tested biomarkers as well as the clinical and PET-based parameters.

### ***Combination of blood biomarkers with clinical and PET data***

The biomarker selection was then added to the parameters that were predictive for response in an externally validated prediction model based on clinical and PET-scan data.<sup>9</sup> The final model consisted of eight variables: tumor length, clinical T stage, clinical N stage, CEA, IL-6, IL-8, osteopontin, and maximal SUV on PET before start of treatment. In the current dataset tumor length was not a significant predictor for response to chemoradiation, but cT and cN were. Maximal SUV was only predictive for response, not for pCR. In Figure 5.1 the ROC curves for the combined models based on biomarkers and clinical data (Figure 5.1A and 5.1A\*) and biomarkers, clinical data and PET information (Figure 5.1B and 5.1B\*) are depicted, as well as the resulting AUCs Figure 5.1C and 5.1C\*). The model based on biomarkers only resulted in an AUC that was more or less comparable to the clinical model.



**Figure 5.2** Nomogram for response prediction. A score for each predictor can be read out at the top scale (Score). All summed scores (Sum of scores scale) can be converted directly to the probability of response.

The AUC of the clinical model was 0.64 (95% CI 0.56-0.71) for pCR and 0.66 (95% CI 0.60-0.72) for response. Combining clinical parameters and biomarkers (AUC 0.69 (95% CI 0.62-0.77) for pCR and 0.73 (95% CI 0.65-0.79) for response) made the model stronger than the model based on biomarkers only or clinical data only. This effect was most pronounced for prediction of response. Addition of information obtained from a PET-scan acquired before the start of treatment lead to strongest models for the prediction of both pCR and response, resulting in AUCs of 0.81 (95% CI 0.73 - 0.88) for pCR and 0.78 (95% CI 0.71-0.85) for response.

The data of the combined prediction model were used to build a hypothesis generating nomogram, which is depicted in Figure 5.2.

## Discussion

To the best of our knowledge this is the first prospective study evaluating the predictive value of a broad range of blood biomarkers analyzed in a standardized way for response prediction to CRT in rectal cancer. CEA turned out to be a significant predictor for pCR and CEA and IL-8 were predictive for response. Including these blood biomarkers in models based on clinical parameters and PET-based parameters resulted in an increased performance of the prediction models.

Osteopontin is a glycoposphoprotein secreted by different cell types, that has been shown to be associated not only with malignancies but also with acute and chronic inflammatory processes. It influences adhesion, migration, invasion, chemotaxis and cell survival<sup>16</sup> and is a marker of tumor aggressiveness and early progression<sup>17</sup>. Although the role of osteopontin as a prognostic factor has been studied extensively in a broad range of solid tumors, only one study looked at the predictive value of osteopontin levels and response to CRT in rectal cancer. Debucquoy *et al.* found an association between lower osteopontin levels and better response in 30 rectal cancer patients, which is in line with our findings, but this did not reach statistical significance.<sup>13</sup> In our patient group it was a significant predictor in univariate analysis, but it lost significance in multivariate analysis. This could be explained by the fact that osteopontin and IL-8 showed a positive correlation.

CEA is an antigen produced by the normal fetus and only in very low concentrations by normal cells of the adult body. In the tumor environment it plays a role in intercellular recognition and attachment. It has been shown to be of prognostic value in colorectal cancer independent of clinical stage and differentiation grade.<sup>18</sup> Although preoperative CEA levels cannot be used to make treatment decisions in colorectal cancer, they consistently have a prognostic value. Patients having a CEA level >5 µg/l have a

significant worse disease free and overall survival than patients with lower CEA levels. Three studies looked at the predictive value of CEA levels for response to CRT in rectal cancer. Das *et al.* studied a group of 562 rectal cancer patients treated with CRT.<sup>7</sup> In their patient group CEA was only predictive for pCR in univariate analysis. However, they chose to dichotomize CEA levels (below and above 2.5 µg/l), while in our model CEA was incorporated as a continuous variable. Yoon *et al.* did an analysis in a group of 351 rectal cancer patients.<sup>12</sup> In multivariate analysis CEA levels ≤5 µg/l were predictive for downstaging and complete regression. Park *et al.* did a retrospective analysis in 352 rectal cancer patients.<sup>11</sup> CEA was analyzed as a continuous variable and had a significant predictive value for responders vs. non-responders. A higher CEA was an independent predictor of poor response to CRT and a worse disease free survival.

IL-8 is a pro-inflammatory chemokine that plays a role in attracting neutrophils. Through the activation of phosphatidylinositol-3-kinase (PI3-K) and phospholipase C it can activate the Akt/mTOR and Raf/Mek/Erk pathways, leading to the promotion of angiogenesis, proliferation and survival and the migration of cancer cells.<sup>19</sup> Polymorphisms in IL-8 have been described to be related to an increased risk of recurrence in rectal cancer<sup>20</sup> and the risk of nodal involvement,<sup>21</sup> indicating a possible relationship with tumor biology of IL-8 in rectal cancer. However, until now there are no reports of a potential value of IL-8 in the prediction of response to CRT.

In this study response was measured in two ways: pCR and ypT<sub>0-2</sub>N<sub>0</sub>. Of these 2 endpoints pCR is the most robust, although the definition of pCR can be difficult. In this cohort routine pathological examination was performed and reported in a standardized way. A recent comparison between routine pathological examination and additional step sections in resection specimens showing no viable tumor cells at initial examination, showed no differences in outcome.<sup>22</sup> Furthermore, pooled analysis of a large series of patients included in different studies, showed a clear prognostic value of pCR after CRT for long-term outcome, even if pooled from different studies, indicating that pCR as scored in routine pathology procedures is a valuable endpoint.<sup>2</sup> The most frequently used method to distinguish responders from non-responders is by means of a tumor regression grade (TRG). However, for this cohort TRG was not scored routinely and we chose to use ypT<sub>0-2</sub>N<sub>0</sub> as definition of response. It makes sense to predict the group of good responders, because they could be suitable for less invasive surgery, like transanal resection or TEM-surgery.

Blood biomarkers give information about tumor biology in an indirect way. A more direct insight can be gained from genetic alterations within the tumor. An overview of these molecular biomarkers is given in the review of Kuremsky *et al.*<sup>23</sup> A possible problem related to molecular biomarkers is the heterogeneity in tumors, making it necessary to collect a representative sampling of tumor material and the invasive

procedure needed to collect material. Blood biomarkers have the advantage that they are easy to collect and that they provide information about the “average” tumor. This makes blood biomarkers useful for daily practice. However translation of the results of this study should be done cautiously, because all biomarkers in this study were measured in a standardized way in one single laboratory, using the same kits. Less thorough and sensitive procedures might influence negatively the added value of biomarkers.

Ideally response prediction takes place before the start of a treatment, so that a patient can be offered the treatment with the highest success rate. In the case of response prediction for rectal cancer, clinical factors and pre-treatment PET-scan have been shown to have predictive value before the start of treatment,<sup>9</sup> but the performance of predictive models based exclusively on pre-treatment data typically lies between 0.65 and 0.70. The predictive value of biomarkers only is in the same range, but the combination of biomarkers and other pre-treatment data results in a stronger prediction model.

The data presented here can be seen as a proof of principle that biomarkers contain predictive information for rectal cancer, but external validation of this prediction model is necessary for a better estimation of the performance and reproducibility of the model.<sup>24</sup> For use in clinical practice a stronger performance is desirable. A possibility to strengthen the predictive model is to incorporate response data obtained early during CRT. If this time point lies early in the treatment course, it is still possible to modify treatment. For PETCT it has already been shown that changes in glucose metabolism after 2 weeks of CRT resulted in a stronger prediction model.<sup>8</sup> An intriguing question is whether early changes in blood biomarker levels during CRT can further enhance the performance of this model. This question will be subject of future research.

In conclusion, pre-treatment CEA levels help to predict pCR after CRT for rectal cancer and CEA and IL-8 levels are helpful in the prediction of response to CRT. These blood biomarkers have added value to earlier published prediction models based on pre-treatment clinical- and PET-data and can be used in decision support systems for tailored therapy.<sup>25</sup>

### ***Acknowledgements***

We acknowledge financial support from the CTMM framework (AIRFORCE project, n° 030-103), EU 7th framework program (METOXIA, EURECA), euroCAT (IVA Interreg, [www.eurocat.info](http://www.eurocat.info)), Radiomics (NIH, USA), EU IMI program (QuIC-ConCePT), NIH-QIN (Radiomics of NSCLC U01 CA143062) and the Dutch Cancer Society (KWF UM 2011-5020, KWF UM 2009-4454).

## References

1. Valentini V, Aristei C, Glimelius B, et al. Multidisciplinary Rectal Cancer Management: 2nd European Rectal Cancer Consensus Conference (EURECA-CC2). *Radiother Oncol* 2009; 92:148-63
2. Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 2010; 11:835-44
3. Zorcolo L, Rosman AS, Restivo A, et al. Complete Pathologic Response after Combined Modality Treatment for Rectal Cancer and Long-Term Survival: A Meta-Analysis. *Ann Surg Oncol* 2012;
4. Bokkerink GM, de Graaf EJ, Punt CJ, et al. The CARTS study: Chemoradiation therapy for rectal cancer in the distal rectum followed by organ-sparing transanal endoscopic microsurgery. *BMC Surg* 2011; 11:34
5. Maas M, Beets-Tan RG, Lambregts DM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011; 29:4633-40
6. Glynne-Jones R, Hughes R. Critical appraisal of the 'wait and see' approach in rectal cancer for clinical complete responders after chemoradiation. *Br J Surg* 2012; 99:897-909
7. Das P, Skibber JM, Rodriguez-Bigas MA, et al. Predictors of tumor response and downstaging in patients who receive preoperative chemoradiation for rectal cancer. *Cancer* 2007; 109:1750-5
8. Janssen MH, Ollers MC, van Stiphout RG, et al. PET-based treatment response evaluation in rectal cancer: prediction and validation. *Int J Radiat Oncol Biol Phys* 2012; 82:871-6
9. van Stiphout RG, Lammering G, Buijsen J, et al. Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential PET-CT imaging. *Radiother Oncol* 2011; 98:126-33
10. Lu K, Zhu Y, Sheng L, et al. Serum fibrinogen level predicts the therapeutic response and prognosis in patients with locally advanced rectal cancer. *Hepatogastroenterology* 2011; 58:1507-10
11. Park JW, Lim SB, Kim DY, et al. Carcinoembryonic antigen as a predictor of pathologic response and a prognostic factor in locally advanced rectal cancer patients treated with preoperative chemoradiotherapy and surgery. *Int J Radiat Oncol Biol Phys* 2009; 74:810-7
12. Yoon SM, Kim DY, Kim TH, et al. Clinical parameters predicting pathologic tumor response after preoperative chemoradiotherapy for rectal cancer. *Int J Radiat Oncol Biol Phys* 2007; 69:1167-72
13. Debucquoy A, Goethals L, Geboes K, et al. Molecular responses of rectal cancer to preoperative chemoradiation. *Radiother Oncol* 2006; 80:172-7
14. Dehing-Oberije C, Aerts H, Yu S, et al. Development and validation of a prognostic model using blood biomarker information for prediction of survival of non-small-cell lung cancer patients treated with combined chemotherapy and radiation or radiotherapy alone (NCT00181519, NCT00573040, and NCT00572325). *Int J Radiat Oncol Biol Phys* 2011; 81:360-8
15. Buijsen J, van den Bogaard J, Janssen MH, et al. FDG-PET provides the best correlation with the tumor specimen compared to MRI and CT in rectal cancer. *Radiother Oncol* 2011; 98:270-6
16. Anborgh PH, Mutrie JC, Tuck AB, et al. Role of the metastasis-promoting protein osteopontin in the tumour microenvironment. *J Cell Mol Med* 2010; 14:2037-44
17. Weber GF, Lett GS, Haubein NC. Osteopontin is a marker for cancer aggressiveness and patient survival. *Br J Cancer* 2010; 103:861-9
18. Goldstein MJ, Mitchell EP. Carcinoembryonic antigen in the staging and follow-up of patients with colorectal cancer. *Cancer Invest* 2005; 23:338-51
19. Waugh DJ, Wilson C. The interleukin-8 pathway in cancer. *Clin Cancer Res* 2008; 14:6735-41

20. Gordon MA, Gil J, Lu B, et al. Genomic profiling associated with recurrence in patients with rectal cancer treated with chemoradiation. *Pharmacogenomics* 2006; 7:67-88
21. Kalady MF, Coffey JC, DeJulius K, et al. High-throughput arrays identify distinct genetic profiles associated with lymph node involvement in rectal cancer. *Dis Colon Rectum* 2012; 55:628-39
22. Park SY, Chang HJ, Kim DY, et al. Is step section necessary for determination of complete pathological response in rectal cancer patients treated with preoperative chemoradiotherapy? *Histopathology* 2011; 59:650-9
23. Kuremsky JG, Tepper JE, McLeod HL. Biomarkers for response to neoadjuvant chemoradiation for rectal cancer. *Int J Radiat Oncol Biol Phys* 2009; 74:673-88
24. Lambin P, van Stiphout RG, Starmans MH, et al. Predicting outcomes in radiation oncology--multifactorial decision support systems. *Nat Rev Clin Oncol* 2012;
25. Lambin P, van Stiphout RG, Starmans MH, et al. Predicting outcomes in radiation oncology--multifactorial decision support systems. *Nat Rev Clin Oncol* 2012; 10:27-40



# Chapter 6

## 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

Ruud van Stiphout\*, Vincenzo Valentini\*, Guido Lammering, Maria Antonietta Gambacorta, Maria Cristina Barba, Marek Bebenek, Franck Bonnetain, Jean-Francois Bosset, Krzysztof Bujko, Luca Cionini, Jean-Pierre Gerard, Claus Rödel, Aldo Sainato, Rolf Sauer, Bruce Minsky, Laurence Collette, Philippe Lambin

Published in: Journal of Clinical Oncology 2011; 29(23):3163-72

## Abstract

### *Purpose*

The purpose of this study was to develop accurate models and nomograms to predict local recurrence, distant metastases and survival for patients with locally advanced rectal cancer treated with long-course chemoradiotherapy (CRT) followed by surgery and to allow for a selection of patients who may benefit most from postoperative adjuvant chemotherapy and close follow-up.

### *Methods and materials*

All data (N = 2795) from the five major European clinical trials for rectal cancer were pooled and used to perform an extensive survival analysis and to develop multivariate nomograms based on Cox regression. Data from one trial was used as an external validation set. The variables used in the analysis were sex, age, clinical tumor stage, tumor location, radiotherapy dose, concurrent and adjuvant chemotherapy, surgery procedure and pTNM-stage. Model performance was evaluated by the concordance index (c-index). Risk group stratification was proposed for the nomograms.

### *Results*

The nomograms are able to predict events with a c-index for external validation of local recurrence (LR; 0.68), distant metastases (DM; 0.73) and overall survival (OS; 0.70). Pathological staging is essential for accurate prediction of long-term outcome. Both preoperative CRT and adjuvant chemotherapy have an added value when predicting LR, DM and OS rates. The stratification in risk groups allows significant distinction between Kaplan–Meier curves for outcome.

### *Conclusions*

The easy-to-use nomograms can predict LR, DM and OS over a 5-year period after surgery. They may be used as decision support tools in future trials by using the three defined risk groups to select patients for post-operative chemotherapy and close follow-up (<http://www.predictcancer.org>).

## Introduction

The improvement in locoregional control reported in recent randomized studies resulting from the combination of preoperative chemoradiotherapy (CRT) and standardized surgical technique (total mesorectal excision, TME) in locally advanced resectable rectal cancer has not been associated with improved survival.<sup>1-5</sup> In these trials, the incidence of distant metastases (DM) was 24%-28%, although local recurrence (LR) rates were only 7%-10%. New treatment strategies must therefore include focus on distant control.

Randomized trials showed that certain patient subsets benefit from postoperative adjuvant chemotherapy after preoperative CRT, but the total patient group showed no reduction in DM or improvement in survival.<sup>1,6</sup> A recent European Consensus Conference failed to reach consensus about the benefit of postoperative chemotherapy after CRT because of insufficient evidence.<sup>7</sup>

To improve survival of patients with locally advanced rectal cancer (LARC), a reduction of the current 25% DM rate is needed. However, treating all patients with postoperative chemotherapy means potential exposure to unnecessary chemotherapy for the remaining 75%.<sup>7</sup> An accurate predictive nomogram would therefore be greatly beneficial. Our primary goal was to develop a nomogram that would predict distant metastasis rate and survival for patients with LARC who are treated by neoadjuvant long-course CRT, and to help select patients who may benefit from the addition of postoperative adjuvant systemic treatment.

The optimal follow-up recommendations after radical resection for colorectal cancer are still undefined, with little level 1 evidence.<sup>7</sup> Many cohort and case-control studies have supported the effectiveness of follow-up, but few randomized controlled trials have correlated follow-up and cancer mortality.<sup>8,9</sup> Identifying subgroups of patients at different risks for LR and DM can help identify the appropriate timing and imaging techniques in a more individualized fashion. Thus, our secondary goal is to develop a nomogram for local recurrence and distant metastases to aid selection of follow-up type and intensity based on an individual's risk of both relapse types.

There are many studies in the literature in which the development of nomograms leads to a successful application for oncology prognostics. Nomograms for predicting follow-up outcome for colorectal cancer are scarce.<sup>10-12</sup> Providing nomograms based on five randomized trials, also allows classification of patients according to risk group to define treatment arms in future randomized trials.

## Methods and materials

### Study population

The mathematical models were developed by using data (N = 2795) from five large European randomized clinical trials for LARC (Table 6.1). These trials evaluated different treatment schemes with similar inclusion criteria and similar accrual times (1992-2003). All patients received preoperative radiotherapy (RT, 34–39 days of treatment time on average; 1.8 Gy per fraction) and surgery (36–47 days after the end of radiotherapy on average) and the fluorouracil (FU)-based adjuvant chemotherapy was the same in all patients who received it. There was heterogeneity for RT dose, surgery procedure and the use of preoperative and postoperative chemotherapy (chemo). Metastases at diagnosis were an exclusion criterion for all datasets. Patients with local excision were also excluded because the group was too small (24 patients) to include statistically.

**Table 6.1** Characteristics of the five European randomized trials (N = 3458) for training the prediction models. \*: Only the pre-operative long-course radiotherapy arm was used for this study. Abbreviations: WHO: World Health Organization performance index, arv: anorectal verge.

Name	Study design	Inclusion criteria	Accrual	Pts	Ref
EORTC	<ul style="list-style-type: none"> <li>• Preop. RT</li> <li>• Preop. CRT</li> <li>• Preop. RT + postop. chemo</li> <li>• Preop. CRT + postop. chemo</li> </ul>	<ul style="list-style-type: none"> <li>• T<sub>3</sub> or resect.T<sub>4</sub>M<sub>0</sub></li> <li>• No history of cancer</li> <li>• Age &lt;80</li> <li>• WHO 0–1</li> <li>• Tumor within 15cm arv</li> </ul>	'93–'03	1011	<sup>1</sup>
French	<ul style="list-style-type: none"> <li>• Preop. RT</li> <li>• Preop. CRT</li> </ul>	<ul style="list-style-type: none"> <li>• T<sub>3</sub>,T<sub>4</sub>M<sub>0</sub></li> <li>• Age&lt;75</li> <li>• WHO 0–1</li> </ul>	'93–'03	733	<sup>3</sup>
German	<ul style="list-style-type: none"> <li>• Preop. CRT</li> <li>• Postop. CRT</li> </ul>	<ul style="list-style-type: none"> <li>• cT<sub>3-4</sub>, or cN+</li> <li>• No history of cancer</li> <li>• Age&lt;75</li> <li>• Tumor within 16cm arv</li> </ul>	'95–'02	823*	<sup>18</sup>
Polish	<ul style="list-style-type: none"> <li>• Preop. RT (5x5 Gy)</li> <li>• Preop. CRT (50.4 Gy)</li> </ul>	<ul style="list-style-type: none"> <li>• cT<sub>3-4</sub>,resectable tumor</li> <li>• Age&lt;75</li> <li>• WHO 0–2</li> <li>• No sphincter involvement</li> </ul>	'99–'02	312*	<sup>2,39</sup>
Italian	<ul style="list-style-type: none"> <li>• Preop. CRT</li> <li>• Preop. CRT + postop. chemo</li> </ul>	<ul style="list-style-type: none"> <li>• cT<sub>3-4</sub></li> <li>• No history of cancer</li> <li>• Age&lt;75</li> <li>• Tumor within 15cm arv</li> </ul>	'92–'01	579	<sup>38</sup>

The variables evaluated at diagnosis include sex, age at the date of random assignment, clinical tumor (cT) stage and tumor location. Location was categorized on the basis of the tumor distance measured from the anorectal verge (arv): low, less than 5 cm; mid, 5 to 10 cm; high: more than 10 cm. Clinical nodal stage (cN) was not included in the multivariate models, because of its absence in certain datasets (69%) and the low reliability of the imaging-based nodal staging.<sup>7</sup> Treatment factors included in the analysis were RT dose (<45 Gy, 45 Gy and >45 Gy), administration of concurrent chemotherapy (yes/no), the surgical procedure (low anterior resection [LAR] and abdominoperineal resection [APR]) and administration of adjuvant chemotherapy (yes/no). Data from pathology included tumor (ypT) and nodal (ypN) stages. Predicted outcomes were the presence of LR, DM and overall survival (OS), accompanied by corresponding time-to-event in months. LR was defined as tumor in the pelvis or perineum, or in the anastomosis as diagnosed by histology. DM was defined as evidence of a tumor in any other area given by at least two imaging exams. All causes of death were included.

### ***Statistical analysis***

To compare the contributions of the evaluated factors, all variables were standardized by subtracting the mean and dividing by the standard deviation. Any missing covariate values (for each covariate less than 10%) were imputed by expectation-maximization (EM) imputation.<sup>13</sup> First, we provided pooled Kaplan-Meier estimates of 5- and 10-year event rates and compare curves for different variable values by using log-rank tests.<sup>14</sup> Secondly, the data were split into a training set and an external validation set. The Italian trial was used for external validation because it is the most recent trial reported. With the ratio of training (N=2242) and validation set size (N=553) it is expected that this split will provide balanced statistical power to find and evaluate the predictors.<sup>15</sup> An additional property of the Italian patient cohort is that all patients received long-course RT and chemotherapy (with optional adjuvant chemotherapy) which is currently the standard treatment for LARC.

The nomograms focus on 5-year event rates, as well as the reports of the pooled randomized trials, and the majority of the patients were followed during this period (54% to 74%). Multivariate analyses were performed by using the Cox proportional hazards model. Hazard ratios were calculated from the model coefficients. The model performance for predicting outcome was evaluated by calculating the concordance index (c-index), which is a generalization of the area under the curve (AUC) of the receiver operating characteristic (ROC) curve and can be applied to continuous outcome and censored data.<sup>16</sup> The maximum value of the c-index is 1.0, indicating a

perfect discrimination, whereas 0.5 indicates a random chance to correctly discriminate outcome with the model.

Significant predictors for outcome were selected by training the Cox model (training data) and selecting the variables that have a significant effect ( $p < .05$ ). The models were trained again with the found predictors as input. In these reduced models, all predictors were still significantly contributing. All model performances were calculated by using bootstrapping, in which the evaluated dataset was tested 1000 times, and each time, the data were randomly resampled (patients can occur in the dataset more than once). The model output was converted into a nomogram.<sup>17</sup>

Three risk groups were defined on the basis of the estimated probability for each event using thresholds on the basis of clinical expertise from literature. For LR, the thresholds chosen were 8%, representing the value of the LR rates in the CRT arms of European Organisation for Research and Treatment of Cancer (EORTC) 22921<sup>1</sup>, Fédération Francophone de Cancérologie Digestive (FFCD) 9203<sup>3</sup> and CAO/ARO/AIO 94 [Working Group of Surgical Oncology/ Working Group of Radiation Oncology/Working Group of Medical Oncology of the Germany Cancer Society]<sup>18</sup> studies and 20% in the surgery alone arm of the Dutch TME trial for stage III patients.<sup>4</sup> For DM and OS, the probability thresholds were 15% and 25%, based on DM-free survival in a recent pathologic complete response (pCR) pooled analysis<sup>19</sup> and in the CRT randomized studies<sup>1,3,18</sup> respectively. The model was calibrated using the Wald statistic for the calibration slope with the three risk groups as bins on the validation dataset and calibration-in-the-large.<sup>15</sup> The Cox model and nomogram algorithm were implemented by using Matlab version 7.1 (MathWorks, Natick MA) and Kaplan-Meier analyses were performed in SPSS version 15.0 (SPSS, Chicago IL).

## Results

### *Follow-up analysis*

Of the 3458 patients accrued in the five randomized trials (Table 6.1), 2795 had pre-operative long-course RT with or without chemotherapy and were eligible for analysis (clinical metastasis stage of 0; surgery; known follow-up time). Event rates at 5 years of follow-up were 12.9% for LR, 30.8% for DM and 30.4% for OS. Median follow-up times were 52.4 months for LR, 47.4 months for DM and 55.2 months for OS. Table 6.2 (Kaplan–Meier analyses) shows that sex and age have a significant impact only on OS time, although tumor location affects both DM and OS. The variables clinical stage,

**Table 6.2** Event free rates at 5- and 10-years for the complete randomized pooled dataset, stratified for each variable. Significant differences between the stratified Kaplan–Meier curves are indicated by the p-value.

		N	Local control			Distant control			Overall survival		
			5y	10y	p	5y	10y	p	5y	10y	p
<b>Total</b>		2795	87.1	85.7	-	69.2	65.8	-	69.6	57.3	-
<b>Clinical diagnosis</b>	<b>Sex</b>										
	Male	1961	87.4	85.1	.466	67.9	64.8	.138	67.6	54.1	<.001*
	Female	843	88.7	87.0		72.1	68.1		74.1	64.9	
	<b>Age [years]</b>										
	<50	379	84.5	83.9	.170	68.0	64.5	.191	71.9	63.1	<.001*
	50–60	751	86.9	85.9		70.1	66.6		72.9	65.2	
	60–70	1135	88.7	85.4		67.3	64.5		67.0	54.6	
	>70	539	89.8	87.4		72.8	68.3		68.8	46.9	
	<b>Tumor loc.</b>										
	Low	953	86.0	84.8	.127	64.3	60.1	.001*	64.7	52.6	.001*
	Mid	1461	88.2	85.4		70.9	68.1		71.5	60.8	
	High	369	90.8	88.6		75.4	72.5		74.3	55.6	
	<b>cT-stage</b>										
2	18	94.4	94.4	<.001*	65.8	65.8	.134	83.0	83.0	<.001*	
3	2228	89.0	87.0		70.1	67.0		71.3	59.9		
4	275	77.5	77.5		64.3	62.0		56.7	39.2		
<b>cN-stage</b>											
0	309	93.2	92.6	.776	78.4	76.7	.009*	77.6	74.4	.006*	
1+2	544	93.6	89.6		71.5	68.4		73.7	62.1		
<b>cStage</b>											
II	733	94.4	92.3	.003*	75.4	73.3	.007*	76.3	68.7	.001*	
III	57	83.5	83.5		63.2	55.6		64.8	43.4		
<b>Treatments</b>	<b>RT dose[Gy]</b>										
	<45	50	84.3	84.3	.019*	60.3	56.0	.268	53.1	39.2	.001*
	45	2205	86.7	85.4		68.9	65.7		68.3	56.6	
	>45	547	92.2	87.6		70.8	66.8		75.3	61.5	
	<b>Conc.chemo</b>										
	No	868	81.1	80.2	<.001*	66.2	63.2	.033*	66.6	52.2	.019*
	Yes	1933	90.7	88.2		70.5	67.0		70.8	59.2	
<b>Surg. proc.</b>											
LAR	1600	91.6	89.1	<.001*	74.0	70.9	<.001*	76.2	65.0	<.001*	
APR	1071	84.6	83.2		63.3	59.8		63.2	49.0		
<b>Adj. chemo</b>											
No	1209	85.9	84.9	.020*	68.3	65.5	.404	66.1	53.1	<.001*	
Yes	1572	89.2	86.6		70.0	66.3		72.3	60.4		
<b>Pathology</b>	<b>ypT stage</b>										
	0	299	96.7	94.5	<.001*	85.1	80.4	<.001*	82.2	71.5	<.001*
	1+2	1008	93.7	92.0		81.4	78.6		80.6	69.8	
	3	1329	83.0	80.8		57.9	54.7		62.1	48.0	
	4	74	66.2	66.2		50.7	46.1		40.2	35.9	
<b>ypN stage</b>											
0	1879	91.3	89.6	<.001*	79.0	75.9	<.001*	77.9	65.7	<.001*	
1+2	833	80.6	76.8		47.6	44.2		53.1	40.2		

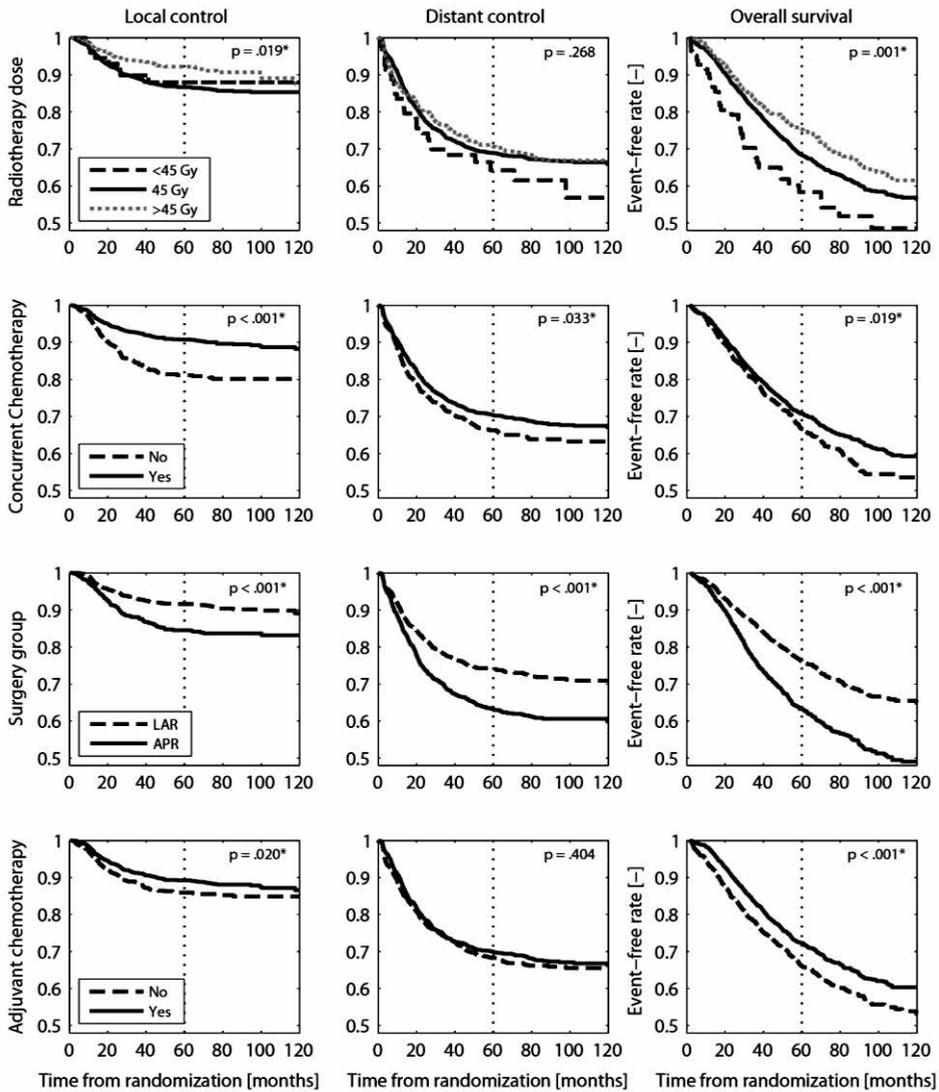
\* Significant overall difference: p<.05

concurrent chemotherapy, surgical procedure, ypT-stage and ypN-stage have significant impact on all outcomes in this analysis. The clinical tumor (cT) stage, RT dose and adjuvant chemotherapy have the same impact, although not for DM. Pathologic staging seems essential in further nomogram development ( $p < .001$  for all outcomes). The effects of all four different treatments on outcome (Figure 6.1) are significant, but RT dose and adjuvant chemotherapy do not have a significant effect on distant metastasis rates.

### ***Nomograms***

For the development of the nomograms, all patients in the training dataset were included ( $N = 2242$ ). Some predictive variables in the Kaplan-Meier analysis (Table 6.2) were not significant anymore in the multivariate models (Table 6.3). Local recurrence prediction performed with a c-index of 0.68, with a 95% confidence interval (CI) for external validation of 0.59–0.76. For distant metastasis, the final model's c-index was 0.73 (95% CI: 0.68–0.77). The nomogram for OS had a c-index of 0.70 (95% CI: 0.65–0.74). These validation performances are not significantly lower than the performances on the training dataset, which makes overfitting by the model less evident.

The resulting nomograms (Figure 6.2) are able to estimate outcome probability by assigning a score (upper scale "Score") to each predictor value. The sum of these scores corresponds to an event probability (bottom two scales). In all nomograms, pathologic staging (ypT and ypN) is the most important factor. Kaplan-Meier estimates of the event rates over time for the validation dataset showed statistically different outcomes for the three proposed risk groups (Figure 6.3);  $p = .010$  for LR,  $p < .001$  for DM and  $p < .001$  for OS. Model calibration was good when applied to the training data, but never significant. In the calibration for external validation it appeared that probabilities were slightly overestimated by the model (LR, 6.4%; DM, 5.7%; OS, 3.7%). Calibration-in-the-large (total average probability) was acceptable for external validation (3.0% - 5.5% probability differences).



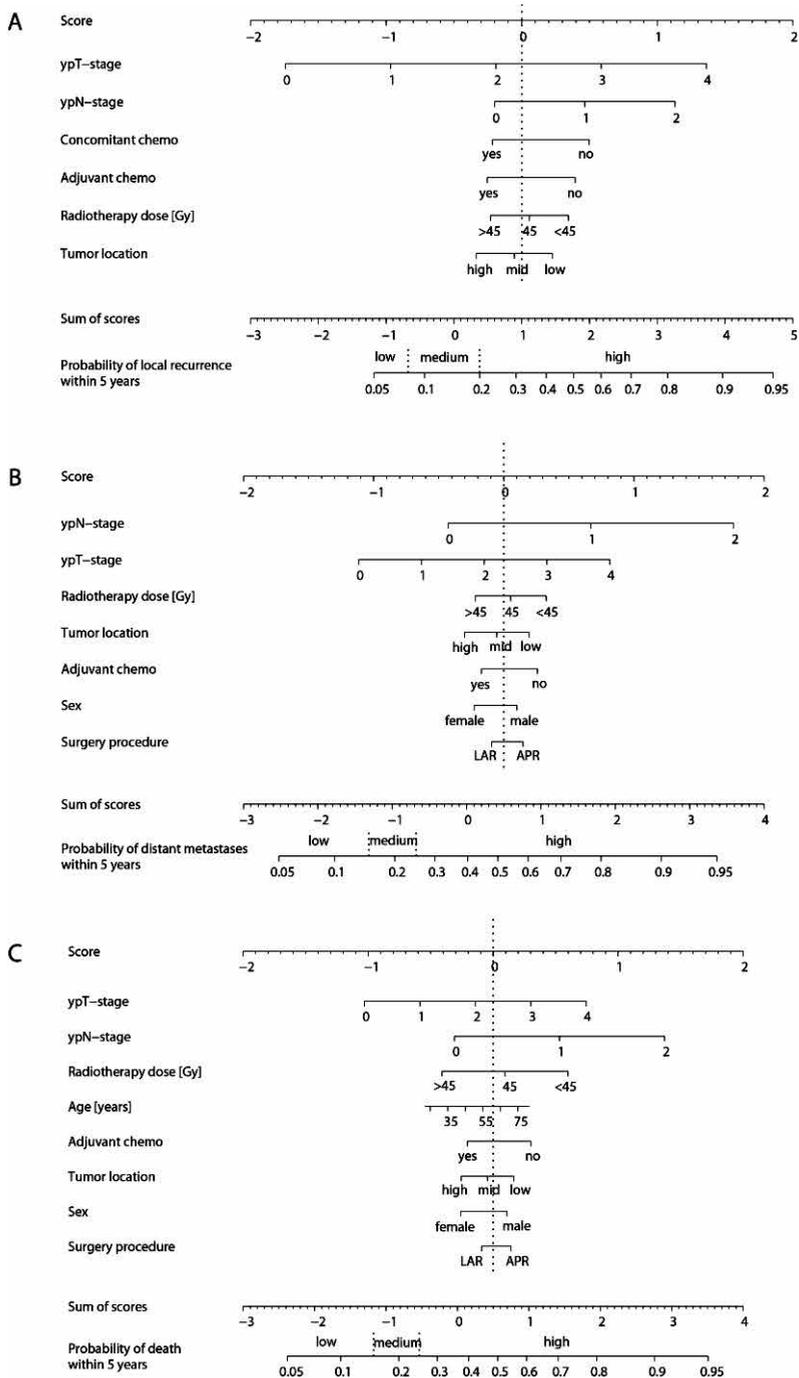
**Figure 6.1** Kaplan-Meier curves stratified for the treatment factors: radiotherapy dose, concurrent chemotherapy, surgery procedure and adjuvant chemotherapy. Exact event rates and p-values are listed in Table 6.2. Dashed line indicates 60 months of follow-up.

**Table 6.3** Model development of five-year outcome: multivariate analyses (Cox proportional hazards regression, hazard ratio HR with confidence intervals and p-value), to determine the final predictors for the nomograms. The c-index (mean and confidence interval) for the training and external validation dataset are given for the derived nomograms as a performance measure.

	Variable	Cox regression			Nomogram
		Hazard ratio (C.I.) <sup>5</sup>	p-value	Performance (c-index)	
Local recurrences	Sex	0.98 (0.87 - 1.10)	.703	Training: 0.71 95% CI: 0.67– 0.74  Validation: 0.68 95% CI: 0.59 – 0.76	
	Age	0.87 (0.78 - 0.97)	.016*		
	cT-stage	1.18 (1.06 - 1.31)	.002*		
	Tumor location	0.97 (0.84 - 1.12)	.644		
	RT dose	0.98 (0.85 - 1.12)	.732		
	Conc. chemo	0.81 (0.72 - 0.91)	.001*		
	Surgery procedure	1.15 (1.00 - 1.33)	.057		
	pT-stage	1.64 (1.38 - 1.95)	<.001*		
	pN-stage	1.26 (1.13 - 1.40)	<.001*		
	Adj. chemo	0.81 (0.72 - 0.92)	.001*		
Distant metastases	Sex	0.94 (0.87 - 1.01)	.100	Training: 0.71 95% CI: 0.69 - 0.74  Validation: 0.73 95% CI: 0.68 - 0.77	
	Age	1.00 (0.93 - 1.08)	.941		
	cT-stage	0.99 (0.91 - 1.06)	.723		
	Tumor location	0.92 (0.84 - 1.00)	.062		
	RT dose	0.95 (0.88 - 1.03)	.193		
	Conc. chemo	1.05 (0.97 - 1.13)	.238		
	Surgery procedure	1.12 (1.03 - 1.23)	.010*		
	pT-stage	1.42 (1.28 - 1.57)	<.001*		
	pN-stage	1.54 (1.44 - 1.64)	<.001*		
	Adj. chemo	0.90 (0.83 - 0.97)	.006*		
Overall survival	Sex	0.87 (0.81 - 0.94)	.001*	Training: 0.68 95% CI: 0.66 – 0.71  Validation: 0.70 95% CI: 0.65 – 0.74	
	Age	1.13 (1.05 - 1.22)	.001*		
	cT-stage	1.13 (1.06 - 1.21)	<.001*		
	Tumor location	0.98 (0.90 - 1.07)	.710		
	RT dose	0.91 (0.85 - 0.98)	.016*		
	Conc. chemo	1.04 (0.97 - 1.12)	.302		
	Surgery procedure	1.18 (1.08 - 1.28)	<.001*		
	pT-stage	1.33 (1.21 - 1.46)	<.001*		
	pN-stage	1.35 (1.26 - 1.44)	<.001*		
	Adj. chemo	0.82 (0.76 - 0.88)	<.001*		

<sup>5</sup> Hazard ratios are calculated in the following framework: sex (female vs male), cT-stage (cT<sub>4</sub> → cT<sub>12</sub>), tumor location (high → low), radiotherapy dose (>45 Gy → < 45 Gy), concurrent chemotherapy (yes vs no), surgery procedure (APR vs LAR), pT-stage (pT<sub>4</sub> → pT<sub>0</sub>), pN-stage (pN<sub>2</sub> → pN<sub>0</sub>), adjuvant chemotherapy (yes vs no). Age is continuous.

\* p-value < .05, selected as predictor for nomogram



**Figure 6.2** Nomograms developed for five year prediction of (A) local recurrence, (B) distant metastases, (C) overall survival. Each variable value is assigned a score, and the sum of scores is converted to a probability in the lowest scale. Calculated probabilities are assigned to a risk group (low, medium, high).

## Discussion

We have developed prediction models for LR, DM and OS for LARC patients based on clinical, treatment and pathological data. Our analysis is based on individual patient data accrued in five European randomized trials that tested preoperative CRT against preoperative RT or postoperative CRT and adjuvant chemotherapy. By pooling the long-course preoperative CRT/RT arms, heterogeneity in the data allowed the models to support decision making in daily practice. However, further external validation of our nomograms by using contemporary patient series is needed.

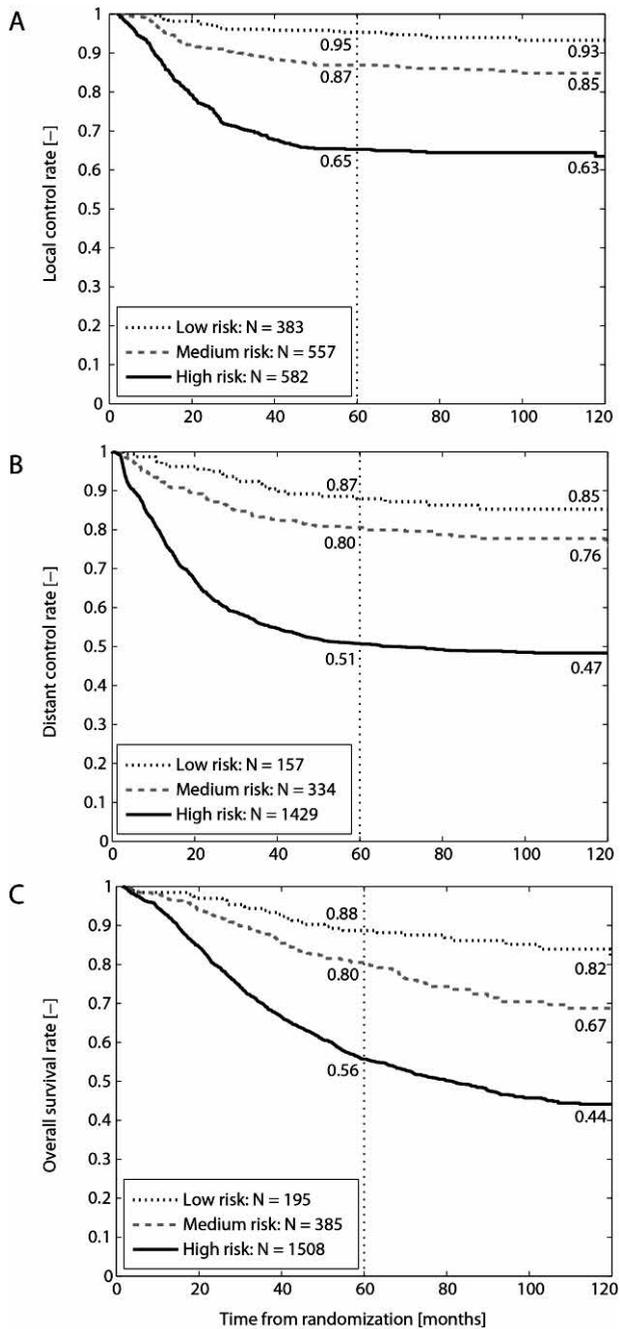
### *Distant metastases prediction*

Our primary goal was to develop a nomogram to predict metastasis rates and select patients for postoperative adjuvant chemotherapy. Adding adjuvant chemotherapy to our pooled analysis improved the DM figures in the metastasis nomogram, but the benefit is small. The best model for DM resulted in the following predictors: ypN-stage, ypT-stage, surgery procedure, and adjuvant chemotherapy (in order of relevance).

According to many studies, patients who achieve a pathologic complete response (ypT<sub>0</sub>N<sub>0</sub>) after preoperative CRT have improved local control rates and improved metastases-free survival rates.<sup>19-21</sup> However, single randomized trials fail to confirm this, possibly because of lack of statistical power.<sup>1,2</sup> Our analysis shows that responding patients had a lower risk of developing metastases.

Pathological studies of the circumferential resection margins (CRMs) at the level of the anorectal junction and anal canal show a high risk of tumor involvement. The quality of surgery in the levator/anal canal area below the mesorectum varies between surgeons, because they may operate in different surgical planes.<sup>22</sup> The higher risk of a residual tumor after an APR related to the more distal location of these cancers could partly explain the impact of the surgical procedure and tumor location in our nomogram. Prospective studies on the reliability of the levator plane (cylindrical APR) in reducing positive CRMs are currently under clinical evaluation.<sup>23-25</sup>

Although the estimated probabilities by the nomograms are on a continuous scale, we proposed three risk groups of DM. Different treatment strategies could be followed for each of these categories (wait and see, 5FU-based regimens and combination of more drugs), and tailored schedule could be randomly tested for any category to test the value of multidrug schedules.



**Figure 6.3** Kaplan–Meier curves of risk group stratification for A. Local control, B. distant control and C. overall survival for validation dataset (N=793). All curves are statistically different (pooled log-rank test,  $\alpha = 0.05$ ).

### ***Overall survival prediction***

Survival is more dependent on the incidence of DM (24% - 28%) than LR rates (7% - 10%). The most predictive model for survival was based on predictors similar of those for LR and DM: ypT-stage and ypN-stage were most relevant, followed by cT-stage, age, adjuvant chemo, surgery procedure, RT dose and sex. Although 50% - 90% of cT<sub>4</sub> patients will be able to undergo a resection with negative margins, depending on the degree of tumor fixation, many still develop a LR as was observed in the univariate analysis and in different phase II studies.<sup>26-28</sup> The selected randomized trials did not report the CRMs and the mesorectal fascia involvement before treatment, so the cT<sub>3</sub> stage could not be stratified. The few cT<sub>1-2</sub> patients were accrued in the randomized trials for the evidence of involved nodes at the staging work-up.

Biologic responses and compliance to cancer treatment seem to change with age: elderly patients showed a worse survival than younger patients, even if the LR nomogram supported less aggressive presentation, with lower LR rates.<sup>29</sup> In any case, even if the randomized studies did not primarily select elderly patients, they demonstrated less favorable compliance with respect to survival when compared with patients in good physical and mental condition, as was also confirmed recently (short-course preoperative RT and TME).<sup>30</sup>

The RT dose affected survival and was related to a better local control (Table 6.2). The association between higher RT dose and better local control and worse sphincter function has been well recognized, although the combination of different radiation doses with concomitant chemotherapy means this topic continues to be debated.<sup>31</sup>

Better survival in females was observed in a postoperative trial of the National Surgical Adjuvant Breast and Bowel Project (NSABP) group<sup>32</sup> some time ago, but it barely reached significance in any subsequent randomized preoperative and postoperative randomized trials. In this analysis, survival associated more favorably with the female gender. Understanding of this difference is worthy of further investigation.

### ***Local recurrence prediction***

Our secondary goal was to develop a nomogram for LR to aid selection of follow-up type and intensity. A recent consensus conference recommended the selection of patients for individualized intensified follow-up, on the basis of age and tumor stage.<sup>7</sup> The best predictive model for LR was based on ypT-stage, cT-stage, age, ypN-stage, concomitant and adjuvant chemotherapy.

Pathological downstaging remains the most relevant predictor of a low risk of LR. Concomitant chemotherapy has confirmed its contribution to ameliorating local control rates as shown in the EORTC 22921<sup>1</sup> and FFCD 9203<sup>3</sup> randomized trials.

Pending the long-term outcome of oxaliplatin-based randomized trials<sup>33,34</sup> at present, infusional 5FU and oral fluoropyrimidines remain the standard agents to combine with preoperative RT.<sup>7</sup>

5FU-based adjuvant chemotherapy reduced the risk of LR in the randomized dataset. To improve survival, the effect of this modality on LR incidence seems to be more important than for DM (Figure 6.1). This observation was also reported by a recent multicenter retrospective analysis.<sup>35</sup> Thus, the nomograms can help clinicians recommend adjuvant chemotherapy according to the improvement in local control rather than for DM prevention.

The nomograms allow identification of the contribution of adjuvant chemotherapy to the estimated outcome risk for any patient. When a responder patient (ypT<sub>0</sub>N<sub>0</sub>) of age 65 years after LAR is compared with a non-responder patient (ypT<sub>3</sub>N<sub>1</sub>) after APR, the nomogram shows the small contribution of adjuvant chemotherapy for responders: For LR, the risk decreases from 3% to 2% for responders and from 20% to 12% for non-responders. For DM, the risk decrease is 11% to 9% for responders and 51% to 36% for non-responders. For OS, the risk decreases from 11.5% to 9.5% for responders and 68% to 57% for non-responders.

Comparing our nomograms to the nomogram for recurrence in colon cancer from the Memorial Sloan-Kettering Institute,<sup>10</sup> reveals that the number of nodes ratio is found as the most important factor in the colon nomogram, which was only recently proposed for rectal cancer.<sup>36</sup> The addition of preoperative carcinoembryonic antigen (CEA) level is predictive, as reported earlier for rectal cancer.<sup>37</sup> Despite the fact that the colon nomogram is based on retrospective data and is not validated, sample size is significant, and the model's performance of 0.76 shows that these factors may be useful for further model development. The other nomogram by Massacesi et al<sup>12</sup> predicting survival, also confirms the importance of tumor response and, in addition, has CEA levels, number of metastatic sites and overall performance status.

### **Statistical analysis**

The proposed nomograms have reliable c-indices for external validation (LR, 0.68; DM, 0.73; OS, 0.70). This performance is good for clinical assistance, but is still not optimal. Adding other knowledge domains to the prediction model, such as traditional and molecular imaging, tumor biology and patient immune competence, is expected to increase model accuracy. Access to population-based registries for further validation will also increase model reliability. At present, the models represent a way to manage

subgroup-level evidence from the available randomized trials by number and to support the design of tailored studies according to different risk categories.

This analysis did not overcome all uncertainties related to surgical quality control, treatment heterogeneity, missing values, unknown treatment compliance, and the unknown salvage treatment required for any pooled analysis. The trial effect was present for adjuvant chemotherapy; studies<sup>1,38</sup> randomized for adjuvant chemotherapy showed no adjuvant chemotherapy effect on outcome, although the pooled dataset did. However, this may have been caused partly by the increase in number of patients. Future meta-analyses may provide more insight on the trial effect, but to achieve data heterogeneity, pooled data were preferred. The current statistical design does not account for competing risk of death for LR and DM prediction. Therefore, small overestimation of LR and DM risk is expected. However, interpretability and comparability with prognostic literature were motivations to use the framework of Kaplan-Meier method and the Cox proportional hazards model.

In conclusion, our analyses suggest that a set of clinical, treatment and pathological variables after preoperative CRT or RT and surgery allows the development of nomograms to predict LR, DM and OS in patients with LARC. We propose a stratification in three risk groups that facilitates the selection of patients for clinical trials with different treatment approaches. Therefore, the nomograms provide decision support for the delivery of postoperative adjuvant chemotherapy for patients with a predicted risk of DM, LR, and OS and a rationale for more individualized follow-up for patients based on the risk of LR and DM.

### ***Acknowledgements***

- To all the groups who provided data
- The contribution of Laurence Collette to this research was supported by the EORTC Charitable Trust

## References

1. Bosset JF, Collette L, Calais G, et al. Chemotherapy with preoperative radiotherapy in rectal cancer. *N Engl J Med* 2006; 355:1114-23
2. Bujko K, Nowacki MP, Nasierowska-Guttmejer A, et al. Long-term results of a randomized trial comparing preoperative short-course radiotherapy with preoperative conventionally fractionated chemoradiation for rectal cancer. *Br J Surg* 2006; 93:1215-23
3. Gerard JP, Conroy T, Bonnetain F, et al. Preoperative radiotherapy with or without concurrent fluorouracil and leucovorin in T3-4 rectal cancers: results of FFCD 9203. *J Clin Oncol* 2006; 24:4620-5
4. Peeters KC, Marijnen CA, Nagtegaal ID, et al. The TME trial after a median follow-up of 6 years: increased local control but no survival benefit in irradiated patients with resectable rectal carcinoma. *Ann Surg* 2007; 246:693-701
5. Sebag-Montefiore D, Stephens RJ, Steele R, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): a multicentre, randomised trial. *Lancet* 2009; 373:811-20
6. Quasar Collaborative G, Gray R, Barnwell J, et al. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet* 2007; 370:2020-9
7. Valentini V, Aristei C, Glimelius B, et al. Multidisciplinary Rectal Cancer Management: 2nd European Rectal Cancer Consensus Conference (EURECA-CC2). *Radiother Oncol* 2009; 92:148-63
8. Jeffery GM, Hickey BE, Hider P. Follow-up strategies for patients treated for non-metastatic colorectal cancer. *Cochrane Database Syst Rev* 2002:CD002200
9. Rosen M, Chan L, Beart RW, Jr., et al. Follow-up of colorectal cancer: a meta-analysis. *Dis Colon Rectum* 1998; 41:1116-26
10. Weiser MR, Landmann RG, Kattan MW, et al. Individualized prediction of colon cancer recurrence using a nomogram. *J Clin Oncol* 2008; 26:380-5
11. Kattan MW, Gonen M, Jarnagin WR, et al. A nomogram for predicting disease-specific survival after hepatic resection for metastatic colorectal cancer. *Ann Surg* 2008; 247:282-7
12. Massacesi C, Norman A, Price T, et al. A clinical nomogram for predicting long-term survival in advanced colorectal cancer. *Eur J Cancer* 2000; 36:2044-52
13. Schafer JL: *Analysis of incomplete multivariate data*, 1997, Chapman & Hall, London
14. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53:457-81
15. Steyerberg EW: *Clinical Prediction Models*, 2009, Springer, New York
16. Harrell FEJ: *Regression modeling strategies*, 2001, Springer, New York
17. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* 2008; 26:1364-70
18. Sauer R, Becker H, Hohenberger W, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med* 2004; 351:1731-40
19. Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 2010;
20. Capirci C, Valentini V, Cionini L, et al. Prognostic value of pathologic complete response after neoadjuvant therapy in locally advanced rectal cancer: long-term analysis of 566 ypCR patients. *Int J Radiat Oncol Biol Phys* 2008; 72:99-107

21. Vecchio FM, Valentini V, Minsky BD, et al. The relationship of pathologic tumor regression grade (TRG) and outcomes after preoperative therapy in rectal cancer. *Int J Radiat Oncol Biol Phys* 2005; 62:752-60
22. Quirke P, Steele R, Monson J, et al. Effect of the plane of surgery achieved on local recurrence in patients with operable rectal cancer: a prospective study using data from the MRC CR07 and NCIC-CTG CO16 randomised clinical trial. *Lancet* 2009; 373:821-8
23. Holm T, Ljung A, Haggmark T, et al. Extended abdominoperineal resection with gluteus maximus flap reconstruction of the pelvic floor for rectal cancer. *Br J Surg* 2007; 94:232-8
24. Marr R, Birbeck K, Garvican J, et al. The modern abdominoperineal excision: the next challenge after total mesorectal excision. *Ann Surg* 2005; 242:74-82
25. West NP, Finan PJ, Anderin C, et al. Evidence of the oncologic superiority of cylindrical abdominoperineal excision for low rectal cancer. *J Clin Oncol* 2008; 26:3517-22
26. Janjan NA, Crane CN, Feig BW, et al. Prospective trial of preoperative concomitant boost radiotherapy with continuous infusion 5-fluorouracil for locally advanced rectal cancer. *Int J Radiat Oncol Biol Phys* 2000; 47:713-8
27. Mohiuddin M, Winter K, Mitchell E, et al. Randomized phase II study of neoadjuvant combined-modality chemoradiation for distal rectal cancer: Radiation Therapy Oncology Group Trial 0012. *J Clin Oncol* 2006; 24:650-5
28. Myerson RJ, Valentini V, Birnbaum EH, et al. A phase I/II trial of three-dimensionally planned concurrent boost radiotherapy and protracted venous infusion of 5-FU chemotherapy for locally advanced rectal carcinoma. *Int J Radiat Oncol Biol Phys* 2001; 50:1299-308
29. Valentini V, Morganti AG, Luzi S, et al. Is chemoradiation feasible in elderly patients? A study of 17 patients with anorectal carcinoma. *Cancer* 1997; 80:1387-92
30. Rutten HJ, den Dulk M, Lemmens VE, et al. Controversies of total mesorectal excision for rectal cancer in elderly patients. *Lancet Oncol* 2008; 9:494-501
31. Tepper JE, Wang AZ. Improving local control in rectal cancer: radiation sensitizers or radiation dose? *J Clin Oncol* 2010; 28:1623-4
32. Fisher B, Wolmark N, Rockette H, et al. Postoperative adjuvant chemotherapy or radiation therapy for rectal cancer results from NSABP protocol R-01. *J Natl Cancer Inst* 1988; 80:21-9
33. Aschele C, Pinto C, Cordio S, et al. Preoperative fluorouracil (FU)-based chemoradiation with and without weekly oxaliplatin in locally advanced rectal cancer: Pathologic response analysis of the Studio Terapia Adiuvante Retto (STAR)-01 randomized phase III trial. *J Clin Oncol* 2009; 27
34. Gerard JP, Azria D, Gourgou-Bourgade S, et al. Comparison of two neoadjuvant chemoradiotherapy regimens for locally advanced rectal cancer: results of the phase III trial ACCORD 12/0405-Prodige 2. *J Clin Oncol* 2010; 28:1638-44
35. Kusters M, Valentini V, Calvo FA, et al. Results of European pooled analysis of IORT-containing multimodality treatment for locally advanced rectal cancer: adjuvant chemotherapy prevents local recurrence rather than distant metastases. *Ann Oncol* 2009; 21:1279-84
36. Dekker JW, Peeters KC, Putter H, et al. Metastatic lymph node ratio in stage III rectal cancer; prognostic significance in addition to the 7th edition of the TNM classification. *Eur J Surg Oncol* 2010; 36:1180-6
37. Nissan A, Stojadinovic A, Shia J, et al. Predictors of recurrence in patients with T2 and early T3, N0 adenocarcinoma of the rectum treated by surgery alone. *J Clin Oncol* 2006; 24:4078-84
38. Cionini L, Sainato A, De Paoli A, et al. *Radiat Oncol* in press;
39. Bujko K, Nowacki MP, Kepka L, et al. Postoperative complications in patients irradiated preoperatively for rectal cancer: report of a randomised trial comparing short-term radiotherapy vs chemoradiation. *Colorectal Dis* 2005; 7:410-6

# Chapter 7

Two-year disease-free and pathological complete response as clinical intermediate endpoint using pooled data of randomized trials for locally advanced rectal cancer

Ruud van Stiphout\*, Vincenzo Valentini\*, Guido Lammering, Maria Antonietta Gambacorta, Maria Cristina Barba, Marek Bebenek, Franck Bonnetain, Jean-Francois Bosset, Krzysztof Bujko, Luca Cionini, Jean-Pierre Gerard, Claus Rödel, Aldo Sainato, Rolf Sauer, Bruce Minsky, Laurence Collette, Philippe Lambin

Submitted

## Abstract

### *Purpose*

Personalized treatments require early characterization of a rectal cancer patient's sensitivity to treatment. This study has two aims: (1) identify the main patterns of recurrence and response to the treatments (2) evaluate pathologic complete response (pCR) and two-year disease-free survival (2yDFS) as intermediate endpoints for predictive modeling.

### *Methods and materials*

Pooled and treatment subgroup analyses were performed on five large European rectal cancer trials (2795 patients), who all received long-course radiotherapy with or without concomitant and/or adjuvant chemotherapy. The ratio of distant metastasis (DM) and local recurrence (LR) rates was used to identify patient characteristics that increase the risk of recurrences.

### *Results*

The DM/LR ratio decreased to a plateau in the first two years, revealing it to be a critical follow-up period. Three patient groups were identified: 5%-15% had pCR and were disease free after 2 years (good prognosis), 65%-75% had no pCR but were disease free (medium prognosis) and 15%-30% had neither pCR nor 2yDFS (bad prognosis).

### *Conclusions*

Compared with pCR, 2yDFS is a stronger predictor of overall survival. To adapt treatment most efficiently, accurate prediction models should be developed for pCR to select patients for organ preservation and for 2yDFS to select patients for more intensified treatment strategies.

## Introduction

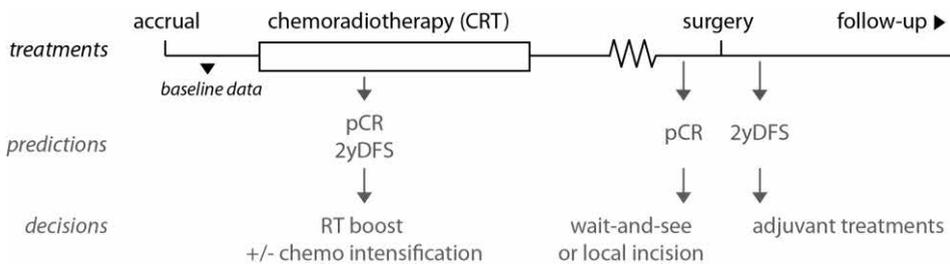
To improve cancer treatment, new hypotheses need to be tested faster in a suitable cohort of patients to avoid diluting treatment for insensible patients. Common methods include the use of large databases<sup>1</sup> and development of reliable surrogate endpoints. A typical example of such an endpoint is the finding that two-year disease-free survival (DFS) is correlated with six-year overall survival (OS) for stage III colon cancer patients treated with adjuvant chemotherapy.<sup>2</sup> Such surrogate endpoints can be identified early and have a high predictive value for survival. By treating them not as prognostic factors but as intermediate outcomes that can be predicted by statistical models, future clinical trials could evaluate individualized treatment strategies, with treatment randomization differentiated according to these predictions. An example of this strategy is model based decision making by using imaging based tumor response predictions, which has potential for radiotherapy boost decisions.<sup>3</sup>

Patients with rectal cancer exhibit heterogeneous responses to treatment; Studies identified a subpopulation of patients who have a medically inoperable disease or who refused surgery but were cured with a low dose of radiotherapy.<sup>4</sup> This subpopulation is characterized by slow growth and long regression time of rectal cancer. Patients who achieved a pCR are considered to be a more favorable subpopulation with less LR and DM as well as better OS.<sup>5,6</sup> While the prediction of early intermediate endpoints (e.g. pCR) is less prone to uncertainties and may therefore be more accurate for individualizing treatment, these endpoints may be weaker prognosticators of long-term outcome.

In this study the value of both DFS and pCR is reported as intermediate endpoints for overall survival in locally advanced rectal cancer patients. The first aim of this study is to identify the optimal time point to assess DFS to best separate the population response-wise. A combination of treatment heterogeneity and a large number of patients in the pooled trial dataset allowed for sub group analyses of recurrence and response patterns. Inspired by Gelman *et al.*,<sup>7</sup> patterns of recurrence are analyzed by calculating the ratio of DM rate and LR rate (which are competing risks) over time to identify a subset of patients with more versus less aggressive disease.

The second aim of the study was to evaluate pathologic complete response (pCR) and two-year disease-free survival (2yDFS) as intermediate endpoints in order to build predictive models which are able to identify patient subgroups before and during treatment. The rate of pCR may be a suitable endpoint to identify a subset of patients who could be successfully treated with more conservative surgery or a non-operative approach. In this setting, prediction of 2yDFS status may be useful to identify

subpopulations with the most aggressive tumors. The two-year timeframe was chosen based on initial inspection of the data, and may allow for selection of different chemotherapeutic regimens to reduce DM rates for this subpopulation by using an accurate prediction for 2yDFS status. If these two intermediate endpoints are able to separate the rectal cancer population in several risk groups accurate models predicting these endpoints may allow for better selection of personalized treatments along the treatment scheme (Figure 7.S1). It is beyond the scope of this paper to deliver accurate prediction models for pCR and 2yDFS, but the focus is on exploration of their role as intermediate endpoints in order to assist future prediction model generation.



**Figure 7.S1** Schematic overview of proposed time points along the treatment to predict pCR and 2yDFS (blue) and their possible corresponding treatment decision (red).

## Methods and materials

### Study population

The analyses were performed on data from five large European randomized clinical trials for locally advanced rectal cancer (N=2795) which were pooled on an individual patient level. This included the EORTC trial<sup>8</sup> comparing preoperative radiotherapy with preoperative chemoradiation and observation with postoperative chemotherapy in a 2x2 factorial design, the French trial<sup>9</sup> comparing preoperative radiotherapy with preoperative chemoradiation, the German trial<sup>10</sup> comparing preoperative chemoradiation with postoperative chemoradiation, the Polish trial<sup>11</sup> comparing preoperative short-course radiation with chemoradiation, and the Italian trial<sup>12</sup> comparing observation with postoperative chemotherapy. The study designs, inclusion criteria, number of patients per trial, and accrual times of the pooled dataset were described in Valentini et al.<sup>13</sup> All patients in the pooled dataset received preoperative radiotherapy (RT) with an average total treatment time of 34-39 days at 1.8 Gy per



fraction, followed by surgery 36–47 days after the end of RT. RT dose and the possible addition of preoperative and postoperative chemotherapy (chemo) introduced treatment heterogeneity into the database. No patients had metastatic disease at the beginning of treatment. A total of 24 patients who underwent a local excision were excluded in advance.

Although many variables were collected in individual clinical trials, this analysis was restricted to assessing the long-term outcome of four subpopulations: 1) those who achieved a pCR and were disease free after 2 years (N=246, 9.2%), 2) those with pCR but also a recurrence (N = 21, 0.8%), 3) no pCR but disease free after 2 years (N = 1868, 70.1%) and 4) no pCR and also a recurrence (N = 528, 19.8%). Additional variables included RT dose (<45 Gy, 45 Gy and >45 Gy), administration of concurrent and/or adjuvant chemotherapy (yes/no), the post-RT tumor localization used to determine the surgery strategy (proximal or distal, depending on the tumor's location from the anal verge with 10 cm cut-off), and the interval between the last fraction of RT and surgery (<28 days, 28-42 days and >42 days).

PCR was defined as the absence of tumor cells in the resected specimen ( $ypT_0N_0$ ). LR was defined as tumor presence in the pelvis or perineum, or in the anastomosis as confirmed by histology. DM was defined as evidence of extrapelvic tumor tissue confirmed by at least two imaging exams. The cumulative incidence of both of these outcomes was reported separately. The absence of both LR and DM within two years was defined as disease free. Patients with less than two years of follow-up time and without occurrence of an event (death or recurrence) were excluded from the DFS analysis (excluded: N=301).

### ***Statistical analysis***

Treatment heterogeneity was accounted for by performing subgroup analyses for the patients receiving different chemotherapy regimens in addition to RT: no chemo, concomitant or adjuvant chemo, or a combination of both chemotherapies. The analyses:

1. The risk ratio at time  $t$  was defined as the ratio of cumulative rate of DM at time  $t$  divided by the cumulative rate of LR at time  $t$ . By calculating the risk ratio for each year during the follow-up period, a time trend of the risk ratio was characterized. Confidence intervals were calculated by bootstrapping the ratio of the two rates from the estimated and normal assumed distributions to arrive at the ratio distribution.
2. The long-term outcome between the non-pCR and pCR populations and between non-2yDFS and 2yDFS, as well as five- and ten-year rates for the cumulative incidence of LR, DM, and OS, were reported<sup>14</sup> and confidence intervals were based on the log-log

transform.<sup>15</sup> All time intervals were measured from the date of randomization, except for the analyses presented in analysis 4 described below.

3. Logistic regression was used to study the impact of characteristics such as sex, age, tumor localization, cT-stage, RT dose, and time to surgery on the odds ratios for the three clinical outcomes of interest: pCR, 2yDFS, and five-year overall survival status (5yOS).

4. The long-term outcomes of the subpopulations defined by their pCR and two-year disease-free status were compared by log-rank test, including the four chemotherapy group stratification. The landmark approach<sup>16</sup> was used to avoid the bias of death within two years affecting the outcome of investigation (2yDFS).

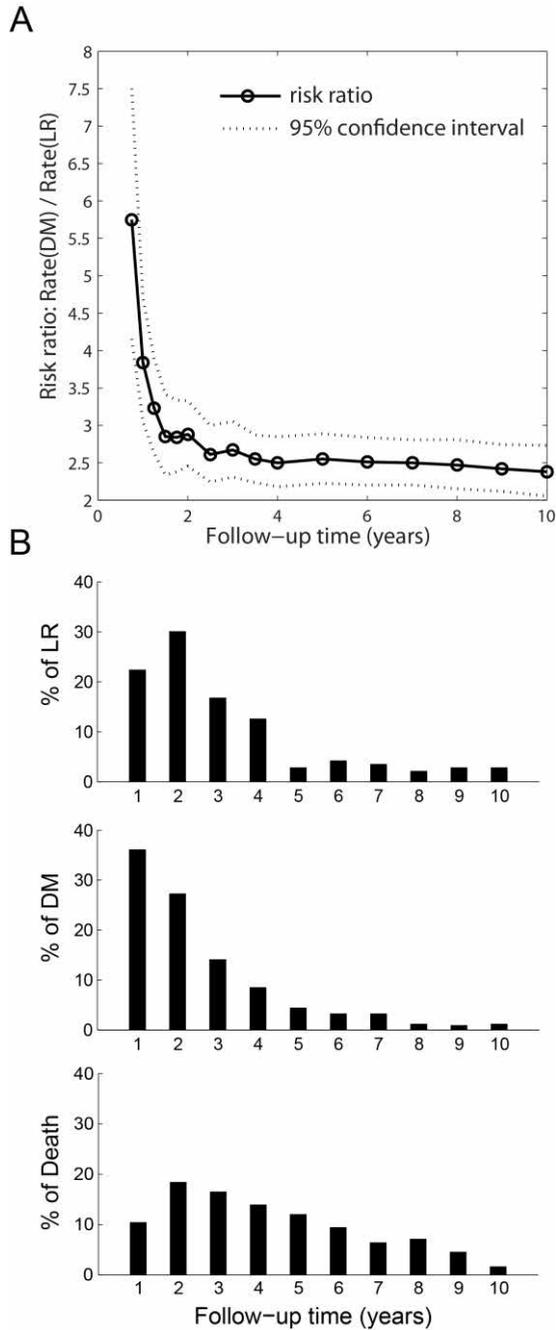
The analyses were conducted in line with the intention-to-treat principle, and performed using SPSS (version 15.0) and MATLAB (version R2010b, MathWorks Inc.).

## Results

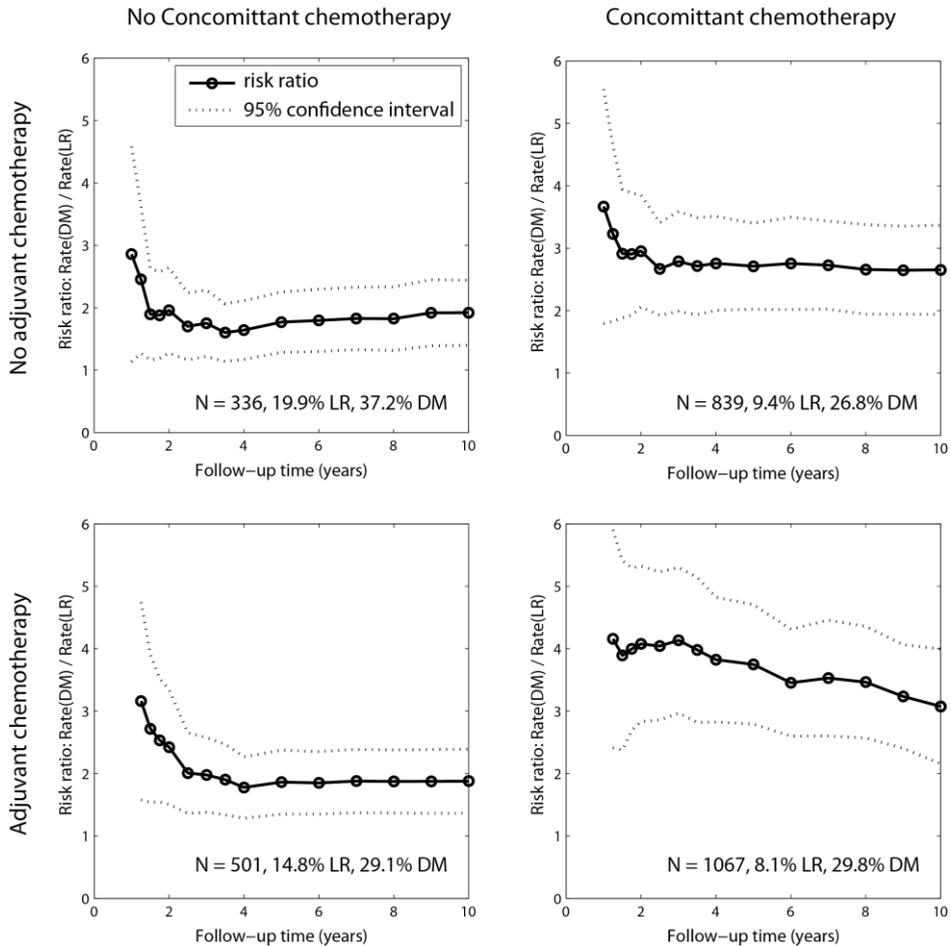
The database characteristics (Table 7.1) are similar in the four subgroups, with the exception of roughly 4% more males in the RT-only subgroup, a higher RT dose in the subgroups with concurrent chemo and both chemotherapies, more pT0 in the concurrent chemo subgroups compared to the RT-only groups (16.7% and 11.2% vs 5.7% and 3.6%, respectively), and roughly 8% more pN<sub>0</sub> for the concurrent chemo subgroup.

### *Risk ratio*

The overall risk ratio between the rates of DM and LR (Figure 7.1A) revealed that, at 1-year follow-up, six times more patients had developed DM than LR (between 4.2 and 7.5, taking 95% confidence intervals into account). This number decreases to approximately 2.5 at two years and remains relatively constant thereafter, with a slight negative slope. Figure 7.1B shows that approximately 80% of the recurrences occur within the first four years, while deaths occur later. For the risk ratios of different chemo treatment schemes (neoadjuvant vs adjuvant), similar curves as in Figure 7.1 are observed (Figure 7.2). These plots show that adding concomitant chemo to RT results in a higher DM/LR risk ratio over time compared to no concomitant chemo, suggesting that concomitant chemo may result in LR reduction without affecting DM incidence (after 2 years concomitant chemo reduces DM by 14.2% and LR by 49.0% compared to the non-concomitant chemo population).



**Figure 7.1** Risk ratio, cumulative rate of distant metastases / cumulative rate of local recurrences, for A. All the patients including 95% confidence interval, and B. Histograms for the percentages of local recurrence and distant metastases occurrences (100%=event occurrence at 10 years) for the total population (N=2795) for each year of follow-up.



**Figure 7.2** Risk ratio, cumulative rate of distant metastases / cumulative rate of local recurrences, for different combinations of given concomitant and adjuvant chemotherapy, including 95% confidence intervals over 10 years of follow-up.

In addition, the population given chemo both adjuvant and neoadjuvant show different behavior on those curves than the other chemo schemes, with the ratio staying relatively constant and decreasing slowly over time. These differences before and after two years are confirmed by the results shown in Figure 7.S2, indicating that aggressiveness of disease is increased for early recurrences, resulting in poor survival, mainly associated with administration of adjuvant chemo. For this reason we consider 2yDFS a reliable intermediate endpoint to stratify the population.

### ***Characterization of populations according to pCR and 2yDFS status***

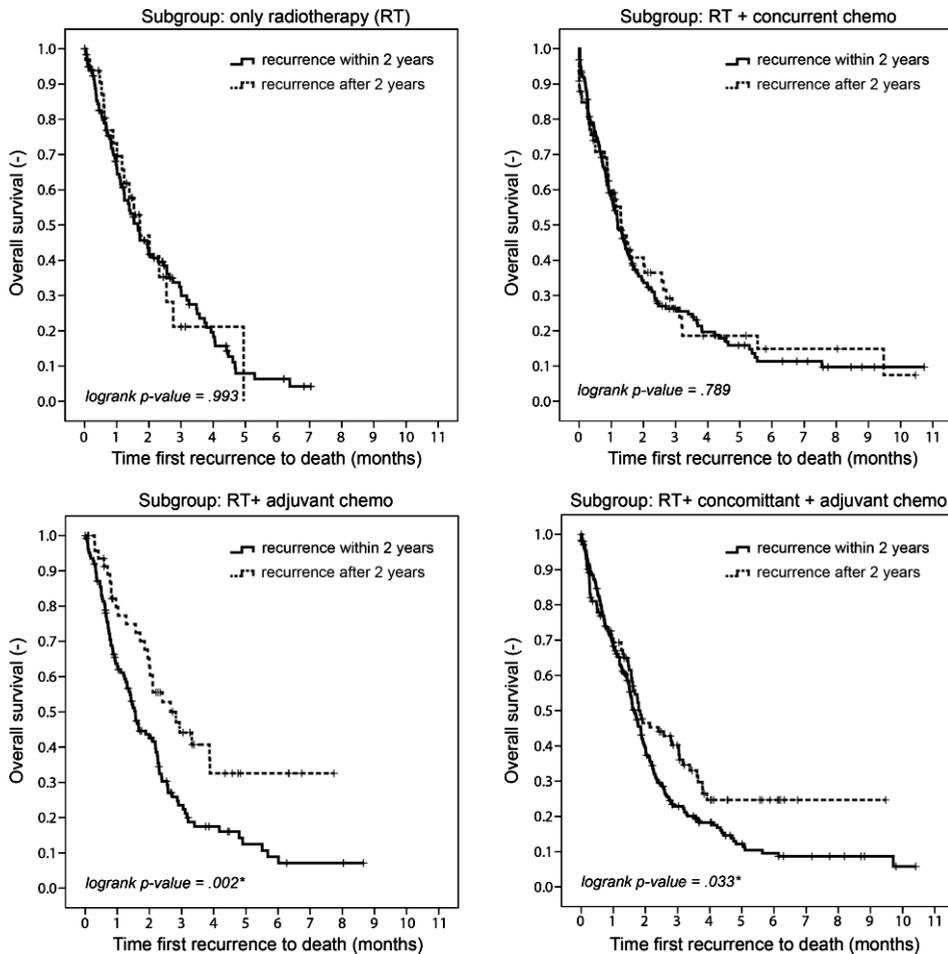
The pCR status was known for 2682 patients; of these, 11.1% had a pCR (N=298). There were significant differences in long-term outcome between the pCR and non-pCR populations (Table 7.2). After ten years, the non-pCR population had a 10.7% higher LR incidence than the pCR population (14.8% vs 4.1%). DM was more frequent in the non-pCR group by 17.1% after ten years (35.2% vs 18.1%). The pCR population had 73.3% overall survivors versus 57% in the non-pCR population (16.6% difference). All differences were significant (log-rank p-values <.001). In the non-2yDFS population 43.1% had LR and 94.8% had DM at five years, and only 12.9% survived. These numbers are far more favorable for the 2yDFS population (6.1%, 16.5%, and 85.3%, respectively).

**Table 7.2** Event rates for pathologic complete response (pCR) vs non-pCR populations and disease free survival at 2 years (2yDFS) vs non-2yDFS at 5 and 10 years of follow-up, including confidence intervals (CI) and p-values for the log-rank test to test if these two populations are significantly different ( $\alpha = .05$ ).

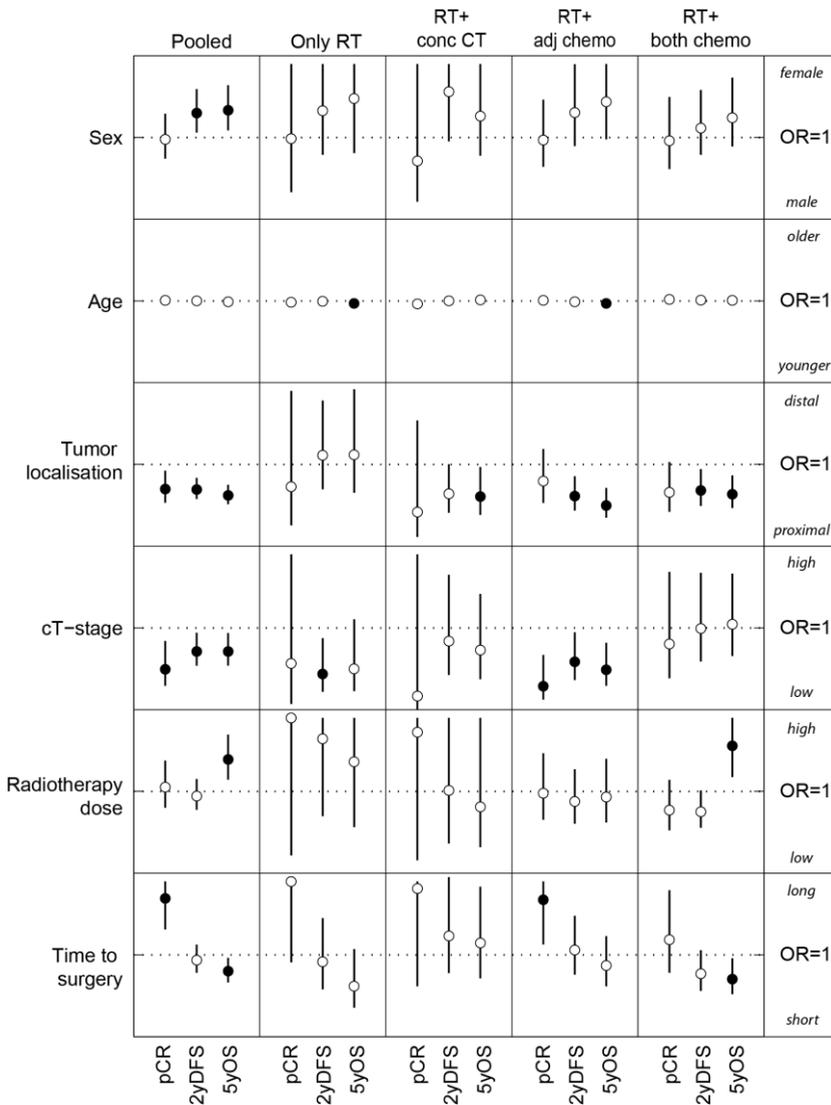
Outcome	N	5 years		10 years		p	HR	[95% CI]
		Rate	[95% CI](%)	Rate	[95% CI] (%)			
<b>LR rate</b>								
Non-pCR	2415	12.8	[11.5-14.2]	14.8	[12.9-16.9]	<.001	0.32	[0.15-0.68]
pCR	267	1.6	[0.6-4.2]	4.1	[1.9-8.7]			
Non-2yDFS	611	43.1	[37.6-49.0]	46.0	[39.1-53.5]	<.001	10.9	[8.5-14.0]
2yDFS	2184	6.1	[5.0-7.4]	8.3	[6.9-10.0]			
<b>DM rate</b>								
Non-pCR	2415	32.3	[30.4-34.3]	35.2	[33.1-37.4]	<.001	0.59	[0.42-0.83]
pCR	267	13.0	[9.3-18.0]	18.1	[12.2-26.5]			
Non-2yDFS	611	94.8	[90.8-97.5]	94.8	[90.8-97.5]	<.001	21.7	[18.3-25.6]
2yDFS	2184	16.5	[14.8-18.4]	20.5	[18.4-22.8]			
<b>OS rate</b>								
Non-pCR	2415	69.5	[71.4-67.5]	57.0	[59.5-54.4]	<.001	0.73	[0.54-0.98]
pCR	267	84.3	[88.4-78.9]	73.3	[80.5-64.0]			
Non-2yDFS	611	12.9	[10.0-16.2]	6.5	[4.1-9.6]	<.001	14.1	[12.2-16.4]
2yDFS	2184	85.3	[83.7-86.8]	71.3	[68.5-73.9]			

### Multivariate analyses

When considering the odds ratios for several predictors (Figure 7.3), 2yDFS better resembles OS than pCR as an outcome, which is significant for sex in the pooled population and for tumor localization in the adjuvant chemo populations. Longer delays between RT and surgery have an increasing effect on pCR rate but not on survival rate.



**Figure 7.S2** Overall survival for recurred patients for two groups with different time periods of the occurrence of first recurrence: within 2 years (N=681) and after 2 years (N=277). Starting point for measuring overall survival is the date of first recurrence.



**Figure 7.3** Odds ratio's (OR) based on the early multivariate prediction of pathologic complete response (pCR), disease free survival at 2 years (2yDFS) and overall survival at 5 years of follow-up (5yOS) with corresponding 95% confidence interval represented by the vertical lines. OR axes are on the same scale for all variables. Six pre-surgical factors were evaluated for their OR: Sex, age, tumor localization, cT-stage, radiotherapy dose and time from last RT fraction to surgery. Beside the pooled database, the four chemotherapy subgroups described in Table 7.1 are also evaluated. Dashed lines describe OR of 1. The direction of the OR is indicated in italic; the direction of the OR indicated a positive effect on outcome (more pCR, more 2yDFS, more 5yOS). White markers represent non-significant difference from 1 and black markers represent a significant different OR from 1.

### ***pCR vs 2yDFS***

Figure 7.4 shows that disease-free status at two years appears to be a better predictor of OS than pCR. The landmark analysis shows that achieving a pCR is significantly beneficial when there is no recurrence within two years of follow-up. In the case of early recurrence, pCR status appears to be statistically irrelevant for OS. The greater effect for OS of achieving 2yDFS than pCR holds for all chemo subgroups. Due to the low number of patients in the subgroup pCR + recurrence within 2 years and the similarity in OS curve development with the no pCR + recurrence group these are considered as one group in the results. This results in three risk groups: 1) pCR + 2yDFS in 246 patients (9.3%, with an interval of 2.7-14.9% depending on treatment group), 2) no PCR + 2yDFS in 1868 patients (70.7% with interval 64.2-76.6%) and 3) no 2yDFS in 528 patients (20% with interval 17.5-29.3%).

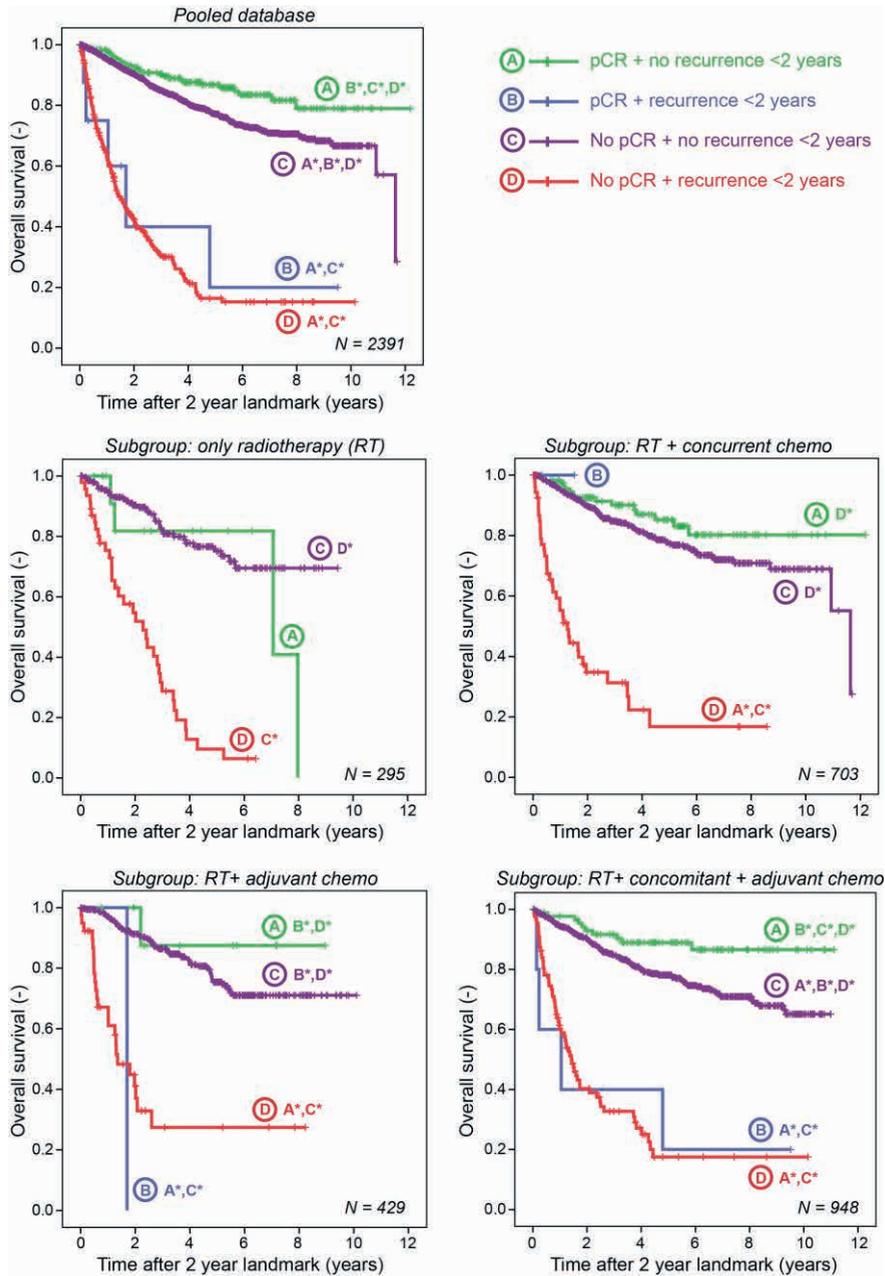
## **Discussion**

This study identified different subpopulations of locally advanced rectal cancer based on their intermediate outcomes in a pooled database of five large European randomized trials. We identified three broad clinical patient subgroups:

- Excellent (~10% with interval 5-15%): curable patients, highly treatable, organ preservation to be pursued, identified by a pCR and 2yDFS.
- Good (~70% with interval 65-75%): curable patients, require effective local treatment, identified by no pCR- and 2yDFS; for some, the role of combined treatment is questionable.
- Poor (~20% with an interval of 17.5-30%): ongoing treatment strategies do not lead to cure or patients having an early and late-occurring recurrent disease; They are identified by having no pCR and no DFS within 2 years.

The overall longer-term aim of the identification of these three identified groups is to develop accurate prediction models for the two intermediate outcomes pCR and 2yDFS with data before and along the treatment. Combined predictions of pCR and 2yDFS (3 risk groups) using clinical, imaging and biomarkers would allow for personalizing treatment.

Variables related to pathological staging after treatment (pT stage and pN stage) were recently found to be the main clinical predictors of LR, DM, and OS,<sup>13</sup> with patients achieving a pCR having an OS benefit of 14.7% at five years and 16.3% at ten years. This is consistent with previous analyses suggesting better outcomes for patients with



**Figure 7.4** Overall survival plots for population stratified for their combined status of pCR and 2yDFS. The time is measured from the chosen landmark of 2 years of follow-up to avoid bias of already deceased patients before that time. The five plots represent the pooled database and the subgroups based on administered concomitant and/or adjuvant chemotherapy. Encircled letters A, B, C and D indicate the represented subgroup. Letters with an asterisk indicate which curves are significantly different from the indicated one with  $\alpha=.05$ .

a pCR.<sup>5,6</sup> These studies reported higher five-year survival rates of 91.6% and 87.6% respectively for patients who achieve a pCR, compared with 84.3% in the present study. However, Maas *et al.*<sup>6</sup> reported a lower survival benefit of 11.2% at five years for patients achieving a pCR. In contrast to our pooled database, both studies used prospective and retrospective results, and almost all patients received concomitant chemo. Therefore, higher absolute rates for the pCR and non-pCR populations are expected. Also, the studies in Maas *et al.* were more recent, and the results may reflect improved treatment regimens and pathology assessment. In this regard, our study shows the long-term benefits of a pCR. In the current study we did not explore formal analyses of surrogacy, as there were too few trials and the treatments were too heterogeneous to perform a valid meta-analysis. Previously, the EORTC 22921 and the FFCD 9203 rectal cancer trials, which are both included in the current pooled dataset, explored the surrogacy of pCR and progression-free survival (PFS),<sup>17</sup> finding that pCR was not a qualified surrogate endpoint for PFS and OS, while PFS was a surrogate endpoint for OS.

For the pCR subgroup with favorable outcomes, lower RT doses or more conservative surgical strategies might be applied to improve their quality of life. To individualize these treatment decisions, reliable prediction models need to be developed to distinguish between these pCR subgroups.

In the pCR group, 61% of the LR cases are reported after five years post treatment (5y LR: 1.6%, 10y LR: 4.1%). To our knowledge, this has not yet been reported in the literature, though it supports Habr-Gama *et al.*<sup>18</sup> These findings indicate that patients with pCR require prolonged observation.

Two-year disease-free status was associated with an OS benefit of 45.5% and 64.8% at five and ten years respectively, compared to patients with recurrence within two years. This supports the identification of a population of patients with aggressive disease that recurs early and leads to poor survival rates, regardless of treatment.<sup>8</sup> While not being the aim of this study, we identified some prognostic factors, such as sex, cT-stage, and tumor localization (Figure 7.3), but more imaging and biological markers should be explored to predict 2yDFS reliably. pCR status is a plausible candidate for 2yDFS based on the studies described above. The comparison between OS according to pCR and 2yDFS showed that the latter had a stronger prognostic impact (Table 7.2, Figure 7.4). A similar result was seen for patients who received either chemoradiotherapy alone or who received both chemoradiotherapy and adjuvant chemo (Figure 7.4).

DFS is superior at predicting survival and should be used as an intermediate endpoint for new prediction models, with the added benefit that the outcome is clear at only

two years follow-up. Accurate prediction of 2yDFS is recommended to select patients for adjuvant chemo regimens.

To decide on more or less aggressive treatments and/or more QoL-oriented treatments, it is important to identify the subgroup of patients with early recurrence, because inclusion in prospective randomized trials may negatively affect the findings. Also, the recurrence analysis of the DM/LR ratios (Figure 7.1) showed a prevalence of metastases in the first two years, indicating a patient group with aggressive disease in whom 5FU based treatment does not appear to reduce the metastasis rate. The not previously reported finding that the DM/LR ratio remained flat after two years, identifies a second group of patients with tumors which can become active after a long follow-up time and have a very slow growth rate. This may be due to an immunological response, and suggests that patients should be followed up for an extended period of time. It appears that for patients receiving concomitant chemo the DM/LR ratio is higher than for the patients without, mainly as a result of the overall reduction of LR induced by concomitant chemo, resulting in improved local effectiveness without corresponding reduction of DM (Figure 7.2). This effect is seen to a lesser extent for DM.

It should be emphasized that, in three trials (N=1244), adjuvant chemo was not randomized or mandated; 73% did receive adjuvant chemo, and the percentage of these patients with pN+ stage was significantly higher than in the group without adjuvant chemo. While this introduces a possible bias, it is likely to be small given the patient numbers involved (14% of the total population had pN+ combined with adjuvant chemo).

### **Conclusions**

This study identified different populations of rectal cancer patients: highly responsive patients (pCR and disease free after 2 years), good responsive curable patients (no pCR but disease free thereafter), and patients with early and late recurrence regardless of pCR status (poor prognosis). While both intermediate outcomes are prognostic (confirming our hypotheses), 2yDFS is stronger than pCR for OS. pCR predictions may identify a minority of highly responsive patients for whom surgery can be tailored, organ preservation promoted, and quality of life or other patient-reported outcomes improved, whereas 2yDFS predictions identify less responsive patients who may benefit from more aggressive chemo. Therefore, 2yDFS should also be considered as an intermediate endpoint in future trials, focusing on identifying and evaluating early predictors for pCR and 2yDFS in order to adapt treatment most efficiently.

## References

1. Sullivan R, Peppercorn J, Sikora K, et al. Delivering affordable cancer care in high-income countries. *Lancet Oncol* 2011; 12:933-80
2. Sargent D, Shi Q, Yothers G, et al. Two or three year disease-free survival (DFS) as a primary end-point in stage III adjuvant colon cancer trials with fluoropyrimidines with or without oxaliplatin or irinotecan: data from 12,676 patients from MOSAIC, X-ACT, PETACC-3, C-06, C-07 and C89803. *Eur J Cancer* 2011; 47:990-6
3. Lambrecht M, Deroose C, Roels S, et al. The use of FDG-PET/CT and diffusion-weighted magnetic resonance imaging for response prediction before, during and after preoperative chemoradiotherapy for rectal cancer. *Acta Oncol* 2010; 49:956-63
4. Wang Y, Cummings B, Catton P, et al. Primary radical external beam radiotherapy of rectal adenocarcinoma: long term outcome of 271 patients. *Radiother Oncol* 2005; 77:126-32
5. Capirci C, Valentini V, Cionini L, et al. Prognostic value of pathologic complete response after neoadjuvant therapy in locally advanced rectal cancer: long-term analysis of 566 ypCR patients. *Int J Radiat Oncol Biol Phys* 2008; 72:99-107
6. Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 2010; 11:835-44
7. Gelman R, Gelber R, Henderson IC, et al. Improved methodology for analyzing local and distant recurrence. *J Clin Oncol* 1990; 8:548-55
8. Bosset JF, Collette L, Calais G, et al. Chemotherapy with preoperative radiotherapy in rectal cancer. *N Engl J Med* 2006; 355:1114-23
9. Gerard JP, Conroy T, Bonnetain F, et al. Preoperative radiotherapy with or without concurrent fluorouracil and leucovorin in T3-4 rectal cancers: results of FFCD 9203. *J Clin Oncol* 2006; 24:4620-5
10. Sauer R, Becker H, Hohenberger W, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med* 2004; 351:1731-40
11. Bujko K, Nowacki MP, Nasierowska-Guttmejer A, et al. Long-term results of a randomized trial comparing preoperative short-course radiotherapy with preoperative conventionally fractionated chemoradiation for rectal cancer. *Br J Surg* 2006; 93:1215-23
12. Cionini L, Sainato A, De Paoli A, et al. *Radiat Oncol* in press;
13. Valentini V, van Stiphout RG, Lammering G, 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. *J Clin Oncol* 2011; 29:3163-72
14. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; 53:457-81
15. Kalbfleisch JD, Prentice RL: *The Statistical Analysis of Failure Time Data* (ed 2), 2002, John Wiley & Sons, Hoboken (NJ) USA
16. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol* 1983; 1:710-9
17. Bonnetain F, Bosset JF, Gerard JP, et al. What is the clinical benefit of preoperative chemoradiotherapy with 5FU/leucovorin for T3-4 rectal cancer in a pooled analysis of EORTC 22921 and FFCD 9203 trials: Surrogacy in question? *Eur J Cancer* 2012; 48:1781-90
18. Habr-Gama A, Perez RO, Proscurshim I, et al. Patterns of failure and survival for nonoperative treatment of stage c0 distal rectal cancer following neoadjuvant chemoradiation therapy. *J Gastrointest Surg* 2006; 10:1319-28; discussion 1328-9



# Chapter 8

General discussion and future perspectives

The main goal of this thesis was to study the prognostic and predictive value of clinical factors, imaging and blood biomarkers for pathologic complete response and follow-up outcome related to recurrences and survival of patients with locally advanced rectal cancer (LARC). Multivariate prediction models were developed, i.e. trained and validated on large databases. These models can be used as clinical tools that allow for treatment modifications in an early stage based on individual predictions. This assistance in decision support requires predictions with high accuracy, which can be only achieved by a holistic approach in which various data sources, multiple time-points and validated statistical methods are used. This chapter summarizes the studies and findings in this thesis and discusses them in a broad perspective and with respect to relevant literature. Furthermore, the implications of this work for the future of individualized medicine are presented.

## Response prediction

The main motivation to assess tumor response to treatment in locally advanced rectal cancer is the ability to tailor treatment for certain risk groups. For example, an early prediction during chemoradiotherapy (CRT) would allow for tumor response improvement by giving a radiotherapy (RT) boost to patients for which a good but not complete response is predicted.<sup>1</sup> For patients with a high chance of a complete response no additional treatment would be required, reducing also unnecessary toxicity. For these complete responder candidates a wait-and-see policy after CRT is a possibility in order to reduce treatment-related morbidity and mortality, for which excellent results are reported.<sup>2</sup> Alternatively, reduced surgery with for example a transanal excision is an option if there is still evidence for a small remaining lesion. These organ sparing approaches increase the quality of life for patients because possible consequences of surgery, like permanent stomas, faecal incontinence and sexual dysfunction, can be avoided. For patients undergoing a wait-and-see approach or reduced surgery an intensified follow-up is however required to track down possible local recurrences as early as possible. Because of the focus on identifying patients with high probability of a complete response, pathologic complete response (pCR) was used as the predicted outcome in all presented response prediction studies. Three sources of data were used for tumor response assessment and prediction in this thesis: pre-treatment clinical data like demographics and tumor stage, <sup>18</sup>F-fluorodeoxyglucose (FDG) positron emission tomography (PET) imaging and blood biomarker data.

### ***Clinical predictors***

When assessing clinical data for response assessment TNM-staging is often firstly considered because of its availability and dominance in making treatment decisions. All patients included in the response prediction studies in this thesis were metastasis free ( $M_0$ ). Clinical tumor (cT) stage and clinical nodal (cN) stage were found to be associated with pCR in all three studies (chapter 3-5) when evaluating only pre-treatment data. A recent analysis with large amount of patients found that cT-stage was predictive for pCR but that cN-stage shows only a trend.<sup>3</sup> This study contained however also old cases where computed tomography (CT) imaging was used for cN-stage scoring. The cT-stage and cN-stage stayed significantly associated with pCR when PET-imaging data before and during treatment was added, but dropped out of the equation when PET information just before surgery was added. It is expected from prediction theory that data measured at this late time point have more impact than the data measured pre-treatment (like cT- and cN-stage). Another finding was that the tumor dimensions are associated with pCR. In all the models the tumor length, assessed by mainly MRI imaging at diagnosis, or the maximum diameter in the metabolic tumor volume (MTV) were predictive for pCR in a multivariate setting. Although tumor length shows a high correlation with tumor volume, the latter was never selected by the models. Others have found that changes in metabolic volume in the pre-post treatment setting were significant for pCR but not the intermediate case (between pre-treatment and during treatment, two weeks after start of CRT).<sup>4</sup> Changes in tumor volume or length over time were however not found to be significant in the presence of PET activity data. This can be explained by the high correlation between this tumor dimension change and reduction in PET activity (chapter 4). However, clinical data pre-treatment alone has low prediction accuracy; area under the curve (AUC) of the receiver operating characteristic (ROC) curve was only 0.61 (N=677, chapter 3) or 0.63 (N=267, chapter 5). For that reason, other sources of data were taken into account.

### ***PET imaging***

Currently, the leading candidate predictive marker for histopathological response prediction in LARC is FDG-PET imaging. A meta-analysis confirmed the added value of PET imaging, especially for intermediate PET imaging (during CRT).<sup>5</sup> The majority of the included studies evaluated pre-CRT versus post-CRT PET imaging and only some assess pre-CRT PET versus intermediate-CRT PET. Both analyses are important for different reasons, as explained above. During CRT, additional treatment decisions like a RT boost can be made based on the available information. Predictions made after CRT are more useful to tailor surgery (reduction or wait-and-see). The decision for a wait-and-see

approach can better be made as late as possible by using both specialized prediction models like the one in chapter 3 and careful assessment of imaging, endoscopies and biopsies.<sup>2</sup> The limitations of the PET studies in the literature are their small sample sizes (numbers of patients are 20-42 for early PET, with a maximal number of patients of 78 for the pre-post setting), the main focus on good versus bad responders (not pCR), the univariate setting in which analyses were performed and the lack of validation. This thesis presented two studies assessing PET imaging in the pre-post setting (chapter 3) and during CRT (chapter 4). These studies aim to increase the clinical applicability of the developed decision making tools by using larger numbers of patients, external validation datasets, and multiple sources of data and focusing on pCR.

For the first model taking into account clinical and PET data pre-CRT and post-CRT, the response index of the pre-treatment tumor length, the relative difference (response index, RI) of maximal standard uptake value (SUV) and the  $SUV_{max}$  value in the tumor just before surgery were predictive. Others have also reported (significant) indications that the response index and post-treatment  $SUV_{max}$  are predictive for response, whereas pre-treatment PET data do not provide sufficient predictive power.<sup>6-8</sup> The second model (pre-CRT + intermediate-CRT) shows similarity in predictor choice for SUV and tumor dimensions: cT-stage, cN-stage,  $\Delta SUV_{mean}$  and intermediate maximal tumor diameter. The change in  $SUV_{mean}$  at the intermediate time point was found predictive in earlier studies for responders versus non-responders,<sup>9-11</sup> but  $SUV_{max}$  was usually stronger when predictions for pCR changes are made.<sup>12,13</sup> In our study the response index of  $SUV_{mean}$  was a stronger predictor than the RI of  $SUV_{max}$ . These two measures are also highly correlated, especially since our PET contouring was semi-automatic and calibrated for both institutes, resulting in less variation in  $SUV_{mean}$  due to contouring.<sup>14</sup>

The significant difference in performances that we reported between the performance of the clinical model and the post-CRT PET data model reflects what others have found in their post-treatment PET analyses. Addition of pre-treatment PET only increases AUC from  $\sim 0.6$  to  $\sim 0.7$ , whereas addition of repeated PET increases performances to acceptable levels; AUC was 0.83 for training and 0.86 for validation for the pre-post model and respectively 1.0 and 0.78 for the pre-intermediate model. The studies that report AUC in general predict good versus no response instead of pCR, which has often higher accuracy.<sup>11,13</sup> Response prediction is in practice more accurate because the number of events for good response (45-55%) is much higher than those of pCR (15-30%). Furthermore, small studies, although very controlled, might be biased towards a

positive outcome by chance. One might also expect that the accuracy of predictions at a later stage (post-CRT) should be more accurate than during CRT, but this is not always the case.<sup>11</sup> The first reason for this can be that the radiotherapy leads in some cases to peritumoral inflammatory tissue, which also avidly consumes FDG.<sup>15</sup> This can result in underestimation of the metabolic response in the post-CRT scan. Measuring inflammation requires for example dynamic FDG-PET imaging, which is discomforting for the patient and more complex and more time-consuming to analyze. Another reason for unreliable assessment of post-CRT PET scans is the partial volume effect (PVE).<sup>16</sup> The PVE could lead to an underestimation of FDG uptake levels, in particular in small tumor volumes. Because CRT shrinks tumors, PVE will affect assessment of post-CRT scans more than early repetitive scans.

There are diverse factors influencing the quantification of PET imaging, like for example the used equipment, scanning protocol, uptake time of the radioactive tracer, the patient blood glucose level (BGL) and the time of scan acquisition (relative to the time of tracer injection). Several factors have been corrected for in the presented studies. Correcting SUV measures for the measured blood glucose level was previously shown to have a positive effect on response prediction.<sup>17</sup> In the presented studies SUV measures were corrected for blood glucose, but most studies do not correct for it, even when it is known that glucose is in competition with FDG in metabolically active tissues.<sup>18</sup> The time between tracer injection and acquisition is also an important factor because after 60 minutes, which is the guideline in most protocols, FDG is still accumulating in the tumor at fast rates. This means that large variations of SUV measures are present for deviations in acquisition time. Scanning after 90 minutes would be better but that will reduce patient throughput on the scanner.<sup>19</sup> In the prospective THUNDER trial (chapter 4) it was difficult to harmonize protocols for e.g. acquisition time and reconstruction algorithms, also because different scanner types in different countries were involved. However, the semi-automatic contouring of the tumor on PETCT was standardized by performing phantom studies in the involved institutes, thereby calibrating the background corrected SUV thresholds.<sup>20</sup> When interpreting multicentric PET studies one is advised to take the above described differences into account and to acknowledge any standardization.

### ***Blood biomarkers***

Blood biomarkers as additional predictors are attractive because the measurements are accurate, standardized, cost-effective and not discomforting for the patient. Previous studies on the predictive value of blood biomarkers are limited to studies mainly evaluating one or two biomarkers. The most studied biomarker for response to

CRT in rectal cancer is carcinoembryonic antigen (CEA).<sup>21-23</sup> Only one other study analyzed the predictive value of osteopontin and interleukin-6.<sup>24</sup> This thesis presented a study in which the added predictive value of 9 pre-treatment blood biomarkers for response was evaluated (chapter 5). The most significant predictor for pCR was CEA, because it remained significant after adding the biomarkers to PET related and clinical data. In this study also good response was predicted ( $\gamma\text{T}_{0-2}\text{N}_0$ ) for which CEA and IL-8 were predictive in the multivariate setting. For CEA its association with response to CRT is reported in LARC, but for IL-8 it is not.<sup>21-23</sup> Including these blood biomarkers in models based on clinical parameters and pre-treatment PET-based parameters resulted in an increased AUC of 0.5 on average, however this increase was not significant. The presented data provided a proof of principle that biomarkers contain predictive information for rectal cancer, however extra time points for PET imaging in combination with (repeated) blood biomarker measurements and external validation of the prediction models are required to substantiate this additive value.

The limitation of the presented response prediction studies is the lack of centralized pathology scoring to determine the predicted outcome  $\gamma\text{pTN}$ -staging. However, a recent comparison between routine pathological examination and additional step sections in resection specimens showing no viable tumor cells at initial examination, showed no differences in outcome.<sup>25</sup> Furthermore, pooled analysis of a large series of patients included in different studies, showed a clear prognostic value of pCR after CRT for long-term outcome, even if pooled from different studies, indicating that pCR as scored in routine pathology procedures is a valuable endpoint.<sup>3</sup>

As suggested, the developed models for pCR can be used for either additional treatment when predictions are made during CRT or to tailor surgery for predictions made post-CRT. Whether the found AUCs between 0.8 and 0.9 are accurate enough for clinical practice depends on the cut-off values for the risk groups. In post-CRT response prediction specifically, the focus should be on high specificity to avoid possible under-treatment (less surgery when surgery is required) rather than overtreatment (standard treatment when less surgery could have been considered). In chapter 4 this was considered for the threshold determination to avoid false positives for pCR. To conclude the discussion on response prediction in this thesis, it was found that repeated PET in combination with clinical factors resulted in the highest accuracy and blood biomarkers are a promising addition for response prediction. To our knowledge these are the first studies for rectal cancer showing that a holistic approach for response prediction is required.

## Predicting long-term follow-up outcomes

### *Five years of follow-up*

In recent trials it was found that locoregional control alone is not sufficient to improve survival and therefore distant control must also be taken into account for new treatment strategies.<sup>26-30</sup> Administration of postoperative chemotherapy for all patients is hazardous for the 75% of the patients that do not require additional chemotherapy to improve survival. Therefore, prediction models for distant metastases (DM) and overall survival (OS) are useful to select patients who may benefit from the addition of postoperative adjuvant systemic treatment. The reason to predict local recurrences (LR) on the other hand, might be very helpful in deciding the appropriate intensity of follow-up. High risk patients require intense follow-up in order to reduce mortality rates.<sup>31,32</sup> Chapter 6 of this thesis presented a high-impact study in which LR, DM and OS after 5 years of follow-up were predicted for a large group of patients who were pooled from 5 large European trials. As described in the response prediction section, variables scored later in the treatment scheme were more predictive as is found also in the follow-up models. The found DM predictors were the yielded pathological nodal-stage (ypN), yielded pathological tumor-stage (ypT), surgery procedure, and adjuvant chemotherapy. This means that response (ypTN), as predicted outcome in chapter 3-5, is now one of the main predictors for further prognosis of the patient. The reason why adjuvant chemotherapy was added as candidate predictor is the ability to calculate its effect for a specific patient. The higher risk of a residual tumor after an abdominoperineal resection (APR) related to the more distal location of these cancers could partly explain the impact of the surgical procedure and tumor location in our nomogram. Prospective studies on the reliability of resecting in different surgical planes for reducing positive CRMs are currently under clinical evaluation.<sup>33-35</sup> The same predictors were found for overall survival but with the addition of cT-stage, age, RT dose and sex. Literature confirmed also these predictors for age<sup>36</sup> and sex<sup>37</sup>, but even though the effect of RT on local control is recognized, the effect of different RT doses (in combination with chemotherapy) remains debated.<sup>38</sup> The nomogram for LR contains the predictors ypT-stage, cT-stage, age, ypN-stage, concomitant and adjuvant chemo. It was also found that a recommendation of adjuvant chemo is even more motivated by improvement in local control rather than for DM prevention, implying that all three nomograms need to be considered when making treatment decisions. In another published nomogram<sup>39</sup> nodal involvement was the most important factor for recurrence, which was only recently proposed for rectal cancer.<sup>40</sup> This nomogram and another one for survival<sup>41</sup>, also included

carcinoembryonic antigen (CEA) levels, resulting in nomogram accuracies of 0.76, which confirms the importance of blood biomarkers once more. In the referred survival model, tumor response was equally dominant for further prognosis as was the case in the presented nomograms.

The proposed nomograms have reliable but suboptimal c-indices for external validation (LR, 0.68; DM, 0.73; OS, 0.70). Uncertainties related to surgical quality control, treatment heterogeneity, missing values, unknown treatment compliance might be contributing to these suboptimal performances. Although the estimated probabilities by the nomograms are on a continuous scale, we proposed three risk groups for the outcomes. These predicted risk groups may assist clinicians in selecting different treatment strategies, especially for the combination of more chemotherapeutic drugs and tailored follow-up schedule.

### ***Two years of follow-up***

By choosing earlier clinical endpoints than 5 or 10 years, new hypotheses can be tested faster which is beneficial for improving cancer treatment. Patients who achieved a pCR are considered to be a more favorable subpopulation with less LR and DM as well as better OS. pCR is expected to be more accurate for individualizing treatment, however this endpoint was found to be a weak prognostic factor of OS. Two-year disease-free survival (2yDFS) is an alternative endpoint having a high correlation with five year survival. In this thesis (chapter 7) this surrogate endpoint was not treated as a prognostic factor for survival but as an intermediate outcome that could potentially be predicted by a statistical model. This chapter explored the role of 2yDFS and pCR as intermediate endpoints in order to assist future prediction model generation. Three subgroups of patients were identified: 1. curable patients with a pCR and 2yDFS (5%-10%, highly treatable, organ preservation to be pursued), 2. curable patients with no pCR but a 2yDFS (65%-75%, requiring effective local treatment), and 3. insensitive patients with no pCR or 2yDFS (17.5%-30%).

Better outcomes for patients with a pCR are reported already in chapter 6 and in recent studies.<sup>3,7</sup> However, pCR did not qualify to be a surrogate endpoint for progression-free survival and overall survival.<sup>42</sup> An interesting and never reported finding (although supported by Habr-Gama *et al.*<sup>43</sup>) is that of the patients in the pCR group with a recurrence, 61% of the recurrences occur after five years of follow-up. This implies that even patients with pCR require prolonged observation. The status of being alive and disease-free after 2 years showed a clear benefit in overall survival and was also a stronger prognostic factor than pCR, even in subanalyses on different chemotherapy schemes. The risk ratio (DM rate/LR rate) over time shows also the clear cut-off point at 2 years of follow-up. In the first two years this ratio was high and

decreasing, implying the occurrence of relatively more metastases, which corresponds to the worst responding group of patients. After two years this risk ratio was approximately flat, which identifies a second group of patients with recurring tumors with a very slow growth rate. In conclusion of this study, pCR predictions may identify a minority of highly responsive patients for whom surgery can be tailored, organ preservation promoted, and quality of life or other patient-reported outcomes improved, whereas 2yDFS predictions identify less responsive patients who may benefit from more aggressive chemo. Therefore, 2yDFS should also be considered as an intermediate endpoint in future trials, focusing on identifying and evaluating early predictors for pCR and 2yDFS in order to adapt treatment most efficiently.

## Clinical impact of prediction models

Despite the high regulation in the oncology field, both treatment and technology are surrounded by uncertain factors. Evidence is primarily generated by the outcomes of randomized clinical trials. This evidence is however often unavailable, inconclusive, valid for only a subgroup of patients, outdated or of insufficient quality. Furthermore, there still exists a translational gap between scientific discoveries and clinical practice. Personalized medicine can exist alongside evidence-based medicine and would allow for fast and efficient testing of new treatment strategies. The presented nomograms in this thesis are incremental steps towards the clinical application of this personalized medicine framework.

In general and also in this thesis, the performance of the developed prediction models is evaluated by discrimination measures, like AUC and c-index, and calibration (i.e. predicted outcome probabilities vs real outcome probabilities). However, when assessing the potential impact of these models in clinical practice, these measures alone are not sufficient. Decision curve analysis can reveal the consequences of using a model in terms of net benefit.<sup>44</sup> To do this, it is required to quantify the harms and benefits of a decision, which is tedious because of the lack of data. Another issue which arises is the selection of the weights for false-positives and false-negatives, which are dependent on the hypothesis and they may even differ from patient to patient. As a first step it is advised to carefully design clinical trials with the purpose of validating the prediction models and assessing their impact. In chapter 4 such a trial was already proposed in which one arm with standard treatment for all (CRT+surgery) is compared to an arm receiving individualized treatment based on the prediction model.

In the meantime, the use of the currently published prediction models can and should be promoted among medical professionals. They are the end-users and they will decide if a prediction model provides sufficient assistance in decision making. The success of the model depends on multiple factors, such as its availability and interactivity, which are expected to increase the acceptability. Even models based on large patient populations, with proper external validation, can fail to be accepted within the healthcare community if the model and its output are not easily interpretable, if there is a lack of opportunity to apply the model or if the clinical usefulness is not assessed. The reason the studies presented in this thesis focused on the nomogram representation of the models, is its high degree of interpretation. However, most insight of the model usefulness is gained by frequent application of models, which can be promoted by making them available on the internet. The most well-known online website with interactive prediction tools is Adjuvant!Online (<http://www.adjuvantonline.com>). This website aims to provide decision support for adjuvant therapy (chemotherapy, hormone therapy, etc.) after surgery for early cancer patients. A prediction website which focuses more on decision support for radiotherapy was previously set up by Maastricht clinic (<http://www.predictcancer.org>). The last issue with the acceptance of these models is the statistical methodology barrier. In traditional statistics there is a subset of methods which are used most frequently for cancer related prognostic and predictive studies. In general, alternative methods are not easily accepted in the medical community, as for example methods from machine learning. These methods, like support vector machines, Bayesian networks and random forests, have the potential to predict the classification of clinical outcomes very accurately, as discussed in chapter 2. It is the responsibility of the medical community to consider these methods in clinical publications, but it is also a responsibility of the model developers to clearly explain the motivation for using a particular method, what inputs are required and how to interpret the output of these methods.

## Future perspectives

The future of clinical decision making for rectal cancer using prediction models is dependent on a few crucial developments. The first step is to improve existing models by mining more sources of data, making more data available for training, validating and updating the models, and achieving higher data quality. Next, the assessment of the models' clinical impact is required, with methods like the decision curve analysis

but certainly also with prediction model based clinical trials, as discussed in the previous section. Third, integration of the prediction models within the hospital IT systems is important for daily use in clinical practice. These aspects will be highlighted for this future outlook; the improvement of models using additional data and the implementation of prediction models within the emerging rapid learning healthcare framework.

### ***Data enrichment for model improvement***

Besides PET-imaging, other functional imaging techniques have shown to be helpful in assessing tumor response. Pre-treatment tumor perfusion or changes in tumor perfusion have been correlated before with pathological tumor response.<sup>13,45</sup> Dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) is most frequently used to measure tumor perfusion, but perfusion-CT (pCT) shows comparable results for perfusion quantification for rectal cancer.<sup>46</sup> Quantifying the compactness, or cellularity of the tumor, to assess the diffusion process within the tumor, can be measured by diffusion-weighted imaging (DWI).<sup>47</sup> Also diffusion-related parameters show high correlation with tumor response after treatment.<sup>48</sup> In combination with PET measures, DWI even reached very high accuracies for a small group of patients.<sup>13</sup> This finding indicates the potential of combining different (functional) imaging techniques. Although DWI is non-invasive, which is an advantage over PET-imaging, the protocols for calibration of DWI measures are not standardized to the same degree as for PET. When extra scans (generated outside the current treatment protocol) from different image modalities are suggested by a certain decision support system, cost-effectiveness studies are required to assess if the benefits in model accuracy outweigh the costs. A less expensive and also promising imaging analysis development is the application of radiomics: the high-throughput extraction of large amounts of image features from radiographic images.<sup>49</sup> Radiomics can be applied to scans from clinical routine like diagnostic and treatment-planning scans, which would increase the amount of extracted data, and thereby evidence.

Alongside imaging, outcome prediction will gain more and more from biological markers in either the blood or in (tumor)-tissues, because of reducing measurement and analysis costs. In chapter 5, initial results of pre-treatment blood biomarkers were presented, but it was argued that measurements at multiple time-points would be more predictive, based on the analogy with the repeated PET findings. A more direct insight can be gained from genetic alterations within the tumor tissue collected with biopsies. A possible problem related to molecular biomarkers is the heterogeneity in tumors, making it necessary to collect a representative sample of tumor material,

requiring an invasive procedure to do so. An overview of these molecular biomarkers for rectal cancer is given in the review of Kuremsky *et al.*<sup>50</sup> Gene expression analyses and genomic profiling through sequencing have also a promising prognostic and predictive value for (colo)rectal cancer. Although several validated prognostic gene expression signatures have been successfully developed, there is little overlap between them.<sup>51</sup> Methodological and technical variances may explain this inconsistency, but it also seems evident that colorectal cancer is an extremely heterogeneous disease at the molecular level. Next-generation sequencing further improves our understanding of the biological mechanisms underlying rectal cancer and may provide additional prognostic information in the near future.<sup>52</sup>

### ***Clinical infrastructure***

The current development of a rapid learning healthcare system may be very effective to achieve clinical decision support in daily practice.<sup>53</sup> In such a system, data is routinely generated through patient care and clinical research and fed into an ever-growing databank or set of coordinated databases. High quality data should be available in real time, simultaneously used to improve clinical care, yield quality measures, and focus on research. This system would also expand the pace and magnitude of evidence generation. Developing a rapid learning healthcare system requires transformation of IT infrastructure, standardization of electronic health records and semantic interoperability between data sharing systems (the ability of any communicating system to share unambiguous meaning). Technical changes are however not sufficient for a successful implementation of the system. First of all, several cultural changes should be induced, like commitment across organizations, community participation for infrastructure development, persuasion of stakeholders to open up information silos, leadership, data governance, patient privacy and lastly that top research journals and grant funding bodies will request open source data sharing. Also, by promoting patient activation and empowerment, patient participation in both decision making and care delivery will be increased.

Clinical decision making using prediction models would benefit greatly from such a rapid learning system. Not only will there be more variety in the data and are higher numbers of patients available to train the classifiers, the classifiers can also be validated across different institutes in a more automated and standardized manner. Also, the frequency of updating the models will increase due to constant monitoring of new data. Because of these promising developments, the medical doctor in the 21<sup>st</sup> century will be assisted by up-to-date tools that can predict multiple patient outcomes with high accuracy with just a single click on a button.

## References

1. Gerard JP, Chapet O, Nemoz C, et al. Improved sphincter preservation in low rectal cancer with high-dose preoperative radiotherapy: the lyon R96-02 randomized trial. *J Clin Oncol* 2004; 22:2404-9
2. Maas M, Beets-Tan RG, Lambregts DM, et al. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. *J Clin Oncol* 2011; 29:4633-40
3. Maas M, Nelemans PJ, Valentini V, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* 2010; 11:835-44
4. Sun W, Xu J, Hu W, et al. The role of sequential 18(F) -FDG PET/CT in predicting tumour response after preoperative chemoradiation for rectal cancer. *Colorectal Dis* 2013; 15:e231-8
5. Zhang C, Tong J, Sun X, et al. 18F-FDG-PET evaluation of treatment response to neo-adjuvant therapy in patients with locally advanced rectal cancer: a meta-analysis. *Int J Cancer* 2012; 131:2604-11
6. Amthauer H, Denecke T, Rau B, et al. Response prediction by FDG-PET after neoadjuvant radiochemotherapy and combined regional hyperthermia of rectal cancer: correlation with endorectal ultrasound and histopathology. *Eur J Nucl Med Mol Imaging* 2004; 31:811-9
7. Capirci C, Valentini V, Cionini L, et al. Prognostic value of pathologic complete response after neoadjuvant therapy in locally advanced rectal cancer: long-term analysis of 566 ypCR patients. *Int J Radiat Oncol Biol Phys* 2008; 72:99-107
8. Konski A, Li T, Sigurdson E, et al. Use of molecular imaging to predict clinical outcome in patients with rectal cancer after preoperative chemotherapy and radiation. *Int J Radiat Oncol Biol Phys* 2009; 74:55-9
9. Avallone A, Aloj L, Caraco C, et al. Early FDG PET response assessment of preoperative radiochemotherapy in locally advanced rectal cancer: correlation with long-term outcome. *Eur J Nucl Med Mol Imaging* 2012; 39:1848-57
10. Cascini GL, Avallone A, Delrio P, et al. 18F-FDG PET is an early predictor of pathologic tumor response to preoperative radiochemotherapy in locally advanced rectal cancer. *J Nucl Med* 2006; 47:1241-8
11. Janssen MH, Ollers MC, van Stiphout RG, et al. PET-based treatment response evaluation in rectal cancer: prediction and validation. *Int J Radiat Oncol Biol Phys* 2012; 82:871-6
12. Goldberg N, Kundel Y, Purim O, et al. Early prediction of histopathological response of rectal tumors after one week of preoperative radiochemotherapy using 18 F-FDG PET-CT imaging. A prospective clinical study. *Radiat Oncol* 2012; 7:124
13. Lambrecht M, Deroose C, Roels S, et al. The use of FDG-PET/CT and diffusion-weighted magnetic resonance imaging for response prediction before, during and after preoperative chemoradiotherapy for rectal cancer. *Acta Oncol* 2010; 49:956-63
14. Buijsen J, van den Bogaard J, van der Weide H, et al. FDG-PET-CT reduces the interobserver variability in rectal tumor delineation. *Radiother Oncol* 2012; 102:371-6
15. Strauss LG. Positron Emission Tomography: Current Role for Diagnosis and Therapy Monitoring in Oncology. *Oncologist* 1997; 2:381-388
16. Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med* 2007; 48:932-45
17. Janssen MH, Ollers MC, van Stiphout RG, et al. Blood glucose level normalization and accurate timing improves the accuracy of PET-based treatment response predictions in rectal cancer. *Radiother Oncol* 2010; 95:203-8

18. Stahl A, Ott K, Schwaiger M, et al. Comparison of different SUV-based methods for monitoring cytotoxic therapy with FDG PET. *Eur J Nucl Med Mol Imaging* 2004; 31:1471-8
19. Janssen MH, Aerts HJ, Ollers MC, et al. Tumor delineation based on time-activity curve differences assessed with dynamic fluorodeoxyglucose positron emission tomography-computed tomography in rectal cancer patients. *Int J Radiat Oncol Biol Phys* 2009; 73:456-65
20. Ollers M, Bosmans G, van Baardwijk A, et al. The integration of PET-CT scans from different hospitals into radiotherapy treatment planning. *Radiother Oncol* 2008; 87:142-6
21. Yoon SM, Kim DY, Kim TH, et al. Clinical parameters predicting pathologic tumor response after preoperative chemoradiotherapy for rectal cancer. *Int J Radiat Oncol Biol Phys* 2007; 69:1167-72
22. Park JW, Lim SB, Kim DY, et al. Carcinoembryonic antigen as a predictor of pathologic response and a prognostic factor in locally advanced rectal cancer patients treated with preoperative chemoradiotherapy and surgery. *Int J Radiat Oncol Biol Phys* 2009; 74:810-7
23. Das P, Skibber JM, Rodriguez-Bigas MA, et al. Predictors of tumor response and downstaging in patients who receive preoperative chemoradiation for rectal cancer. *Cancer* 2007; 109:1750-5
24. Lu K, Zhu Y, Sheng L, et al. Serum fibrinogen level predicts the therapeutic response and prognosis in patients with locally advanced rectal cancer. *Hepatogastroenterology* 2011; 58:1507-10
25. Park SY, Chang HJ, Kim DY, et al. Is step section necessary for determination of complete pathological response in rectal cancer patients treated with preoperative chemoradiotherapy? *Histopathology* 2011; 59:650-9
26. Bosset JF, Collette L, Calais G, et al. Chemotherapy with preoperative radiotherapy in rectal cancer. *N Engl J Med* 2006; 355:1114-23
27. Bujko K, Nowacki MP, Nasierowska-Guttmejer A, et al. Long-term results of a randomized trial comparing preoperative short-course radiotherapy with preoperative conventionally fractionated chemoradiation for rectal cancer. *Br J Surg* 2006; 93:1215-23
28. Gerard JP, Conroy T, Bonnetain F, et al. Preoperative radiotherapy with or without concurrent fluorouracil and leucovorin in T3-4 rectal cancers: results of FFCD 9203. *J Clin Oncol* 2006; 24:4620-5
29. Peeters KC, Marijnen CA, Nagtegaal ID, et al. The TME trial after a median follow-up of 6 years: increased local control but no survival benefit in irradiated patients with resectable rectal carcinoma. *Ann Surg* 2007; 246:693-701
30. Sebag-Montefiore D, Stephens RJ, Steele R, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): a multicentre, randomised trial. *Lancet* 2009; 373:811-20
31. Jeffery GM, Hickey BE, Hider P. Follow-up strategies for patients treated for non-metastatic colorectal cancer. *Cochrane Database Syst Rev* 2008:CD002200
32. Rosen M, Chan L, Beart RW, Jr., et al. Follow-up of colorectal cancer: a meta-analysis. *Dis Colon Rectum* 1998; 41:1116-26
33. Holm T, Ljung A, Haggmark T, et al. Extended abdominoperineal resection with gluteus maximus flap reconstruction of the pelvic floor for rectal cancer. *Br J Surg* 2007; 94:232-8
34. Marr R, Birbeck K, Garvican J, et al. The modern abdominoperineal excision: the next challenge after total mesorectal excision. *Ann Surg* 2005; 242:74-82
35. West NP, Finan PJ, Anderin C, et al. Evidence of the oncologic superiority of cylindrical abdominoperineal excision for low rectal cancer. *J Clin Oncol* 2008; 26:3517-22
36. Rutten HJ, den Dulk M, Lemmens VE, et al. Controversies of total mesorectal excision for rectal cancer in elderly patients. *Lancet Oncol* 2008; 9:494-501
37. Fisher B, Wolmark N, Rockette H, et al. Postoperative adjuvant chemotherapy or radiation therapy for rectal cancer: results from NSABP protocol R-01. *J Natl Cancer Inst* 1988; 80:21-9

38. Tepper JE, Wang AZ. Improving local control in rectal cancer: radiation sensitizers or radiation dose? *J Clin Oncol* 2010; 28:1623-4
39. Weiser MR, Landmann RG, Kattan MW, et al. Individualized prediction of colon cancer recurrence using a nomogram. *J Clin Oncol* 2008; 26:380-5
40. Dekker JW, Peeters KC, Putter H, et al. Metastatic lymph node ratio in stage III rectal cancer; prognostic significance in addition to the 7th edition of the TNM classification. *Eur J Surg Oncol* 2010; 36:1180-6
41. Massaccesi C, Norman A, Price T, et al. A clinical nomogram for predicting long-term survival in advanced colorectal cancer. *Eur J Cancer* 2000; 36:2044-52
42. Bonnetain F, Bosset JF, Gerard JP, et al. What is the clinical benefit of preoperative chemoradiotherapy with 5FU/leucovorin for T3-4 rectal cancer in a pooled analysis of EORTC 22921 and FFCD 9203 trials: Surrogacy in question? *Eur J Cancer* 2012;
43. Habr-Gama A, Perez RO, Proscuschim I, et al. Patterns of failure and survival for nonoperative treatment of stage c0 distal rectal cancer following neoadjuvant chemoradiation therapy. *J Gastrointest Surg* 2006; 10:1319-28; discussion 1328-9
44. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; 26:565-74
45. de Lussanet QG, Backes WH, Griffioen AW, et al. Dynamic contrast-enhanced magnetic resonance imaging of radiation therapy-induced microcirculation changes in rectal cancer. *Int J Radiat Oncol Biol Phys* 2005; 63:1309-15
46. Kierkels RGJ, Backes WH, Janssen MHM, et al. Comparison Between Perfusion Computed Tomography and Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Rectal Cancer. *International Journal of Radiation Oncology\*Biography\*Physics* 2010; 77:400-408
47. Boone D, Taylor SA, Halligan S. Diffusion weighted MRI: overview and implications for rectal cancer management. *Colorectal Dis* 2013; 15:655-61
48. Lambregts DM, Vandecaveye V, Barbaro B, et al. Diffusion-weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study. *Ann Surg Oncol* 2011; 18:2224-31
49. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48:441-6
50. Kuremsky JG, Tepper JE, McLeod HL. Biomarkers for response to neoadjuvant chemoradiation for rectal cancer. *Int J Radiat Oncol Biol Phys* 2009; 74:673-88
51. Shibayama M, Maak M, Nitsche U, et al. Prediction of metastasis and recurrence in colorectal cancer based on gene expression analysis: ready for the clinic? *Cancers (Basel)* 2011; 3:2858-69
52. Kim TM, Lee SH, Chung YJ. Clinical applications of next-generation sequencing in colorectal cancers. *World J Gastroenterol* 2013; 19:6784-6793
53. Abernethy AP, Etheredge LM, Ganz PA, et al. Rapid-learning system for cancer care. *J Clin Oncol* 2010; 28:4268-74



## List of abbreviations

## Biology

<b>CA IX</b>	Carbonic anhydrase 9
<b>CA19-9</b>	Carbohydrate antigen 19-9
<b>CEA</b>	Carcinoembryonic antigen
<b>CRP</b>	C-reactive protein
<b>DNA</b>	Deoxyribonucleic acid
<b>GWAS</b>	Genome-wide association studies
<b>IL-6</b>	Interleukin-6
<b>IL-8</b>	Interleukin-8
<b>LDH</b>	Lactate dehydrogenase
<b>mRNA</b>	Messenger ribonucleic acid
<b>OPN</b>	Osteopontin
<b>SNP</b>	Single nucleotide polymorphism

## Clinical

<b>DFS</b>	Disease free survival
<b>DM</b>	Distant metastasis
<b>LARC</b>	Locally advanced rectal cancer
<b>LR</b>	Local recurrence
<b>NLARC</b>	Non-locally advanced rectal cancer
<b>OS</b>	Overall survival
<b>pCR</b>	Pathological complete response
<b>TRG</b>	Tumor regression grade

## Imaging

<b>5FU</b>	5-fluorouracil
<b>ADC</b>	Apparent diffusion coefficient
<b>CT</b>	Computed tomography
<b>DICOM</b>	Digital imaging and communications in medicine
<b>DW-MRI</b>	Diffusion-weighted magnetic resonance imaging
<b>FDG</b>	18F-fluorodeoxyglucose
<b>MRI</b>	Magnetic resonance imaging

<b>MTV</b>	Metabolic tumor volume
<b>PET</b>	Positron emission tomography
<b>SUV</b>	Standardized uptake value

## Methodology

<b>AUC</b>	Area under the curve
<b>ROC</b>	Receiver operating characteristic
<b>SD</b>	Standard deviation
<b>SVM</b>	Support vector machine
<b>LOO</b>	Leave-one-out
<b>pSVM</b>	Proximal support vector machine
<b>CI</b>	Confidence interval

## Treatment

<b>APR</b>	Abdominoperineal resection
<b>CRM</b>	Circumferential margin
<b>CRT</b>	Chemoradiotherapy
<b>Gy</b>	Gray
<b>IGRT</b>	Image-guided radiotherapy
<b>LAR</b>	Low anterior resection
<b>MRF</b>	Mesorectal fascia
<b>TEM</b>	Transanal endoscopic microsurgery
<b>TME</b>	Total mesorectal excision
<b>APR</b>	Abdominoperineal resection

## Miscellaneous

<b>CDSS</b>	Clinical decision support system
<b>EORTC</b>	European organisation for research and treatment of cancer
<b>FFCD</b>	Fédération francophone de cancérologie digestive
<b>THUNDER</b>	Theragnostic utilities for neoplastic diseases of the rectum
<b>WHO</b>	World health organization



## Summary

## Summary

The epidemiology and the treatment of rectal cancer were briefly discussed in **chapter 1**. Preferably, the response to treatment and the consequence for a patients' prognosis is assessed as early as possible in order to be able to adapt treatment accordingly. This concept of individualized medicine requires the development of specific computational tools to predict outcome after treatment because over the last few years the amount of available medical information has expanded rapidly. The opportunities for the physician to make a detailed assessment of risk and benefits associated with a specific combination of tumor, patient and treatment characteristics are expected to be greatly increased with the use of these tools. Eventually, these tools will be implemented within the infrastructure of hospitals and used in daily care. This thesis provided a first step towards these clinical decision support systems (CDSS) for rectal cancer.

All the studies presented in this thesis focused on locally advanced rectal cancer (LARC), which is often treated with long course chemoradiotherapy (CRT) followed by surgery (TME). For LARC, the options to individualize treatment are highlighted throughout the thesis. One option is to predict pathologic complete response (pCR) in order to identify patients that respond excellent to the CRT. These patients are possible candidates for a wait-and-see policy in the surgical setting. Patients who respond less well might benefit from an extra boost of radiotherapy to promote more pCR cases. The worst responding group can be selected for additional chemotherapy. On the other hand, prediction of long-term follow-up outcomes, generally assessed in a 5-year time frame after treatment, may assist in the planning of follow-up strategies. For example, a high estimated risk for a local recurrence (LR) motivates an intensification of the follow-up, i.e. making the time shorter between hospital visits. High estimated risks for distant metastases (DM) and events for overall survival (OS), on the other hand, promotes the administration of adjuvant chemotherapy.

*The main aim of this thesis was to study the predictive value of clinical factors, imaging and blood biomarkers for response to treatment and follow-up outcome related to recurrences and survival for patients with locally advanced rectal cancer.*

In **chapter 2** a review about predicting outcomes in the field of radiation oncology is presented. An overview of clinical, treatment, imaging and molecular factors that are associated with outcomes in radiation oncology is provided. Furthermore, the methodology behind the development of statistical prediction models is discussed.

This development does not only involve finding the relevant factors that influence outcome but also the training of a statistical model that is able to make predictions for new patients. To train models accurately and to validate them independently, datasets of large populations of representative patients are required. Integration of these validated decision support systems in the clinic that are constantly updated using new data (“rapid learning”) requires harmonization of data infrastructures and the willingness to accept these systems in healthcare organizations, so that data and knowledge are being shared in a standardized, instant and global manner.

### **Tumor response prediction using PETCT-imaging and blood biomarkers**

Prediction of pathological complete response (15-30% of the patients) after CRT using clinical factors and sequential positron emission tomography (PET) combined with computed tomography (CT) imaging was described in **chapter 3**. This study involved large datasets from four different European centers. For the patients with imaging, a PETCT scan was made one week before and another scan around 8 weeks after treatment. It was found that the addition of the post-treatment PETCT scan information to the clinical data and the first PETCT scan resulted in a significant increase in prediction performance with acceptable levels of accuracy with area under the ROC curve (AUC) of around 0.85. The dominant tumor dimension and the maximal uptake of radioactive isotopes in the tumor as well as its relative difference between PET scans were found to be the best predictors for pCR.

The time point at which the actual prediction of pCR is calculated is favored at an earlier stage to be able to adapt treatment. For this reason a large prospective study with PETCT imaging before and during CRT was presented in **chapter 4**. Tumor dimensions and change in PET activity within the tumor, together with clinical stage of the tumor, were found to be predictive for pCR. The model was externally validated and performed well (100% accuracy for the training set and 67% in validation).

We are convinced that a holistic approach of combining multiple sources for data greatly improves the accuracy of prediction models for response to treatment. Blood biomarkers and their relationship with tumor response were studied in **chapter 5** in a prospective study of 276 patients. Pre-treatment carcinoembryonic antigen (CEA) level is a significant predictor for pCR after CRT for LARC. For good response, i.e. yielded pathological tumor stage of 0-2 and no nodal involvement ( $ypT_{0-2}N_0$ ), CEA and IL-8 levels were both predictive. These blood biomarkers also had an added value to the

model with pre-treatment clinical- and PETCT-data. These results suggest that blood biomarkers have predictive potential for tailored therapy in rectal cancer.

### **Predicting long-term follow-up outcome for decision support**

In **chapter 6** a large study with 2795 patients from five European institutes aimed at the development of accurate prediction models for local recurrence, distant metastases and survival, based on only clinical data. The trained models were presented as nomograms, which can assign a new patient to one of the three proposed risk groups. This facilitates the selection of patients for clinical trials with different treatment approaches. External validation of the nomograms resulted in AUCs of 0.68 for LR, 0.73 for DM, and 0.70 for OS. Pathological staging was essential for accurate prediction of long-term outcome.

The intermediate clinical outcomes pCR and 2-year disease free survival (2yDFS) were evaluated in **chapter 7**. Based on these intermediate endpoints which are both prognostic, different populations of rectal cancer patients were identified: highly responsive patients, responsive curable patients, and patients with early and late recurrence regardless of pCR status. pCR predictions may identify a minority of highly responsive patients for whom surgery can be tailored, organ preservation promoted, and quality of life or other patient-reported outcomes improved, whereas 2yDFS predictions identify less responsive patients who may benefit from more aggressive chemotherapy. Therefore, 2yDFS should also be considered as an intermediate endpoint in future trials, focusing on identifying and evaluating early predictors for pCR and 2yDFS in order to adapt treatment most efficiently.

Finally, in **chapter 8** the results presented of this thesis and areas of future research were discussed. The clinical impact of prediction models does not only depends on prediction accuracy, but also on the validation of these models in clinical trials and the dissemination of the acquired knowledge and developed tools. For the future, data enrichment using more advanced imaging techniques and analyses and adding biological markers from for example the genomic domain are expected to improve the quality of outcome predictions even more. If those improved models are integrated into a rapid learning healthcare system with suitable IT infrastructures and seamless interconnection between data systems, personalized treatment can and is expected to be fully embraced.





## Samenvatting

## Samenvatting

In **hoofdstuk 1** wordt de epidemiologie en de behandeling van endeldarmkanker kort beschreven. Bij voorkeur wordt de respons op de behandeling en de consequenties op de prognose van een patiënt in een zo vroeg mogelijk stadium ingeschat, om de behandeling daarop aan te passen. Dit noemt men ook wel een geïndividualiseerde behandeling. Het kiezen van de optimale behandeling voor een patiënt is voor artsen echter steeds ingewikkelder geworden omdat de laatste jaren de hoeveelheid informatie betreffende de patiënt en specifieke kenmerken van de tumor is toegenomen. Ook zal het aantal behandelingsopties en combinaties van behandelingen alleen maar stijgen in de toekomst. Daarom is er een behoefte aan beslissingsmodellen die gebruik maken van de capaciteit van computers om al deze informatie te combineren en te koppelen aan behandelingsuitkomst. Het ontwikkelen van deze modellen is een eerste stap naar een Decision Support systeem, dat in de toekomst volledig geïntegreerd zal zijn in het ziekenhuis. In deze thesis zijn wetenschappelijk gefundeerde voorspellingsmodellen voor endeldarmkanker ontwikkeld.

Alle studies in de thesis richten zich op lokaal gevorderd endeldarmkanker (locally advanced rectal cancer, LARC) dat meestal wordt behandeld met een langdurige chemoradiotherapie (CRT) gevolgd door een operatieve ingreep (TME). De opties om voor LARC de behandeling te individualiseren zijn beschreven in de thesis. Een optie is om pathologische complete respons (pCR) te voorspellen met als doel patiënten te selecteren die uitstekend reageren op de CRT. Deze patiënten zijn mogelijke kandidaten voor een beleid waarin een chirurgische ingreep vermeden kan worden ("wait-and-see") en hebben dan ook geen extra behandeling nodig. Patiënten die echter minder goed reageren, zouden een extra dosis radiotherapie kunnen krijgen met als doel het aantal complete responders te vergroten. Als de respons echter afwezig lijkt te zijn, dan kunnen de betreffende patiënten een extra chemotherapie ondergaan. Daarnaast is het ook mogelijk om voorspellingen te doen over behandelingsuitkomsten gedurende het nazorg traject, waarbij een termijn van 5 jaar klinisch relevant is. Een hoog berekend risico op lokaal recidief (terugkeren van de tumor), bijvoorbeeld, pleit voor een intensievere nazorg. Hoge risico's voor uitzaaiingen (metastasen) of overlijden van de patiënt pleiten voor het voorschrijven van adjuvante chemotherapie.

*Het hoofddoel van deze thesis was het bestuderen van de waarde van klinische factoren, beeldvorming en bloed biomarkers voor het voorspellen van respons op de behandeling en nazorg gerelateerde uitkomsten als recidieven en overleving voor patiënten met lokaal gevorderde endeldarmkanker.*

In **hoofdstuk 2** wordt een algemeen overzicht gegeven over de huidige stand van zaken met betrekking tot het voorspellen van uitkomsten in de radiotherapeutische oncologie. Daarbij worden klinische, behandeling gerelateerde, beeldvorming gerelateerde, en moleculaire factoren bediscussieerd waarvan bekend is dat ze geassocieerd zijn met de behandelingsuitkomst. Ook wordt de methodologie besproken die nodig is om statistische voorspellingsmodellen te ontwikkelen. Deze ontwikkelingen omvatten niet alleen het vinden van relevante factoren die klinische uitkomsten beïnvloeden, maar ook het trainen van een statistisch model dat in staat is om voor nieuwe patiënten een voorspelling te doen. Om deze modellen nauwkeurig te trainen en onafhankelijk te valideren zijn grote representatieve datasets nodig met medische informatie over patiënten en tumor karakteristieken. Het integreren van deze voorspellingsmodellen in de kliniek vraagt om een hoge mate van beschikbaarheid van nieuwe data. Hiervoor zijn de juiste infrastructuren en beleidsmanagement nodig. Eenmaal geïntegreerd, kan data en kennis op een gestandaardiseerde, snelle en wereldwijde wijze gedeeld worden.

### **Tumor respons voorspellingen met PETCT beeldvorming en bloed biomarkers**

Het voorspellen van pathologische complete behandelingsrespons (in 15-30% van de patiënten) met behulp van herhaalde beeldvorming met PETCT (positronemissietomografie in combinatie met computertomografie) is beschreven in **hoofdstuk 3**. Deze studie bevat data uit vier verschillende Europese centra. Voor de patiënten die beeldvorming ondergingen werd 1 scan gemaakt in de week vóór de behandeling en 1 scan 8 weken na de behandeling. De gevonden resultaten laten zien dat de informatie uit de klinische gegevens en de PETCT scan van vóór de behandeling significant verbeteren wanneer PETCT informatie van na de behandeling wordt toegevoegd. Dit leidt tot een nauwkeurigheid van het model met acceptabele waarden voor de AUC van 0.85. De voorspellende factoren voor pCR zijn de tumor grootte, de maximale opname van de PET tracer (FDG) in de tumor en de relatieve afname van deze opname gemeten tussen de twee PETCT scans.

Bij voorkeur is het tijdstip waarop de voorspelling voor pCR wordt gedaan zo vroeg mogelijk in het behandelingstraject om de mogelijkheid te behouden om de

behandeling aan te passen. Om deze reden is een grote prospectieve studie opgezet met PETCT beeldvorming vóór en tijdens CRT die is beschreven in **hoofdstuk 4**. De tumor grootte en de verandering van PET tracer opname in de tumor in combinatie met klinische classificatie waren in deze studie voorspellend voor pCR. Het model werd extern gevalideerd en presteerde goed (100% nauwkeurigheid voor de trainingsdataset en 67% voor externe validatie).

We zijn overtuigd dat een holistische benadering met een combinatie van verschillende data bronnen de accuraatheid van voorspellingsmodellen sterk kan vergroten. Daarom zijn ook bloed biomarkers en de relatie met tumor respons onderzocht in **hoofdstuk 5** in een prospectieve studie met 276 patiënten. Carcino-embryonaal antigeen (CEA) waarden vóór de behandeling hadden een significante associatie met pCR. Voor een goede response (tumor classificatie van 0-2 en geen betrokkenheid van lymfklieren, ypT<sub>0-2</sub>N<sub>0</sub>), waren de bloedwaarden van CEA en IL-8 voorspellend. De bloed biomarkers hadden ook een toegevoegde voorspellende waarde in het model met klinische en PETCT factoren. Deze bevindingen wijzen erop dat bloed biomarkers belangrijke kandidaten zijn om therapie op maat voor endeldarmkanker te optimaliseren.

### **Het voorspellen van uitkomsten tijdens de langdurige nazorg**

In **hoofdstuk 6** is een grote studie beschreven met 2795 patiënten die behandeld zijn in vijf Europese instituten. Op basis van klinische data zijn nauwkeurige voorspellingsmodellen ontwikkeld voor lokaal recidieven, metastasen en overleving. De getrainde modellen zijn gepresenteerd als nomogrammen, die in staat zijn om voor één enkele patiënt een risico groep voor de betreffende uitkomst te berekenen. Zo een nomogram faciliteert ook de selectie van patiënten in klinische trials met verschillende behandelingsopties. Externe validatie van de nomogrammen resulteerde in AUCs van 0.68 voor lokaal recidieven, 0.73 voor metastasen en 0.70 voor overleving. Pathologische tumor en lymfklier classificaties na behandeling hadden hierbij de grootste voorspellende waarde.

De tussentijdse klinische uitkomsten pCR en ziekte vrije overleving binnen twee jaar (2yDFS) zijn geëvalueerd in **hoofdstuk 7**. Deze tussentijdse uitkomsten waren prognostisch in de data en op basis hiervan werden verschillende populaties endeldarmkanker patiënten geïdentificeerd: uitstekend reagerende patiënten, goed reagerende behandelbare patiënten en patiënten met vroeg of laat een recidief, ongeacht de status van pCR. Voorspellingen voor pCR kunnen een minderheid van de

patiënten identificeren voor wie een operatie op maat kan worden gegeven, orgaan preservatie kan worden gepromoot en de kwaliteit van leven kan worden verbeterd. Slecht reagerende patiënten kunnen beter met de voorspellingen voor 2yDFS geïdentificeerd worden en komen in aanmerking voor agressievere chemotherapie. Mede daarom, zou overwogen kunnen worden om 2yDFS als tussentijds eindpunt te evalueren in toekomstige klinische trials met als doel om vroege predictoren voor pCR en 2yDFS te vinden en behandelingsaanpassingen efficiënter maken.

Ten slotte worden in **hoofdstuk 8** de resultaten uit de gehele thesis besproken en met elkaar in verband gebracht. De klinische impact van voorspellingsmodellen hangt niet alleen af van de nauwkeurigheid van voorspellingen maar ook van een betrouwbare validatie in klinische trials en van de verspreiding van de opgedane kennis en modellen. Voor de toekomst wordt verwacht dat voorspellingsmodellen zullen verbeteren door het toevoegen van extra relevante data afkomstig van geavanceerde beeldvormingstechnieken en -analyses maar ook het toevoegen van moleculaire biomarkers uit bijvoorbeeld het genomics domein. Deze verbeterde modellen kunnen uiteindelijk pas volledig omarmd worden als ze geïntegreerd gaan worden in de snel lerende zorg systemen die de juiste IT infrastructuur hebben en zorgen voor een naadloze verbinding tussen de data systemen.



## Acknowledgements / dankwoord

## Acknowledgements / dankwoord

Dit promotieboekje is het resultaat van jaren onderzoek, waar ik met veel plezier op terug kijk. Gedurende deze tijd heb ik enthousiaste mensen ontmoet, heb ik met velen samengewerkt, ben ik geïnspireerd geraakt en heb ik veel vrienden gemaakt. Graag wil ik iedereen bedanken die, in welke vorm dan ook, heeft bijgedragen aan dit boekje. Ik had dit niet voor elkaar kunnen krijgen zonder de hulp en aanmoediging van velen, waarvan ik enkele mensen speciaal wil benoemen.

Beste Philippe, bedankt dat je mijn promotor was. Je zat altijd vol ideeën waardoor mijn projecten snel vorm kregen. Ik waardeer je oplossingsgerichte en effectieve manier van communiceren. Ik ben blij dat ik deel mocht uitmaken van het CAT onderzoek dat onder jouw leiding een flinke ontwikkeling heeft doorgemaakt binnen Maastrou.

Dear Vincenzo, although you were not my official supervisor, I consider you as one of my scientific fathers. We have spent many late night hours together on Skype to interpret exciting results or to review our manuscripts. I have learnt a lot from your great clinical expertise. You gave me the opportunity to gain experience with large clinical datasets and to work together with renowned scientists across Europe. The times that I worked with you in Rome were also pleasant. Grazie mille.

Beste Guido, als co-promotor heb je mij met enthousiasme veel geleerd over de klinische gang van zaken bij Maastrou. Bedankt dat ik altijd bij je binnen kon lopen en dat je structuur aanbracht in mijn onderzoek in onder andere de wekelijkse abdomen meeting.

Beste Eric, in het begin van mijn promotie heb je mij kennis laten maken met de wereld van de machine learning. Bedankt voor alle inspirerende meetings en je persoonlijke betrokkenheid. Mijn onderzoek heeft uiteindelijk een iets medischer karakter gekregen, maar ik heb je adviezen altijd erg gewaardeerd. Beste Jaap, bedankt dat je mij de mogelijkheid hebt gegeven om in Tilburg te kunnen werken en mijn onderzoek daar te kunnen presenteren. Joke, bedankt voor alle ondersteuning en je opgewektheid.

Beste Ralf, bedankt dat je als promotor wilt optreden. De gesprekken die we gehad hebben waren aangenaam en ik waardeer de flexibiliteit van de afdeling Knowledge Engineering om mij vrij te laten in het bepalen van mijn onderzoeksrichting.

Ik wil graag de leden van de beoordelingscommissie (Prof.dr. F. Ramaekers, Dr. G. Beets, Prof.dr. R. Beets-Tan, Prof.dr. F. Mottaghy) bedanken voor hun tijd en energie om mijn proefschrift te beoordelen.

My gratitude to all collaborators I worked with: C.Barba, M.Bebenek, A.Begg, F.Bonnetain, J.F.Bosset, P.Boutros, K.Bujko, C.Capirci, L.Cionini, L.Collette, C.Creutzberg, M.A.Gambacorta, M.Gava, J.P.Gerard, W.van Gijn, A.Giordano, P.Granone, K.Haustermans, M.Lambrecht, E.Meldolesi, B.Minsky, R.Nout, C.Rödel, D.Rubello, D.De Ruysscher, A.Sainato, R.Sauer, P.Slagmolen and others. Laurence, thank you for your excellent statistical input. Christina and Maria Antonietta, thank you for making me feel welcome in Rome. Elisa, thank you for all your efforts in the THUNDER trial. Carien en Remi, ik heb veel geleerd van jullie grote expertise in het klinische onderzoek, bedankt. Willem, bedankt voor de fijne samenwerking.

Vervolgens wil ik al mijn geweldige collega's bij Maastricht bedanken voor hun betrokkenheid en gezelligheid. In het begin van mijn promotie werd ik door mijn kamergenoten gastvrij ontvangen. Bedankt Wouter, Hugo, Steven, Marco, Enrica, Erik, Miguel, Esther, Ramon, Lucas, Katia en Georgy voor deze leuke en leerzame periode in de gezellig drukke kamer. Hugo, je hebt me geïnspireerd om groots te denken als onderzoeker. Steven, bedankt voor de gezellige wijnproefavonden en je gastvrijheid mbt de logeerkamer. Wouter, altijd heerlijk om even met een koffie je kantoor binnen te wandelen en de randverschijnselen van het onderzoek grondig door te nemen. Marco, bedankt dat ik je paranimf mocht zijn. Door jouw PET expertise is mijn enthousiasme voor imaging alleen maar toegenomen.

The last couple of years wouldn't have been the same without this group of great guys. Our diversity remarkably turned out to be a golden combination. Thanks Mark, Patrick, Emmanuel, Guillaume, Sean, Mathieu, Pablo, Leo and Fabien for all the fun we had drinking proper coffee, drinking beers in the Falstaff, Café Zuid, Take One, Zondag, or any other favorite pub, and for the unforgettable trips we made in Europe.

Emmanuel, thank you for being my paranimph. It felt like we had our own room in the back of the big room. We shared many stories, music, videos, sports moments, interesting facts, jokes and expressions in Dutch, Spanish and brabants. I appreciate you for being a social animal and for showing me around in Maas. And by that, I don't mean Maaskantje.

Guillaume, thank you also for being my paranimph. You are very hospitable and I appreciate your taste for the good life. You are always very proactive to bring the guys

together. Also, you inspired me and others with your impressive PhD and by mastering the Dutch language so quickly.

Mark, your place was always a safe haven with a spare matras and a cosy kitchen. Thank you for being so helpful and for distracting me with some of your cool ideas. Paddy, thanks for all the fun we had and for the genuine interest you have in the people around you.

Charlie has left the building, angels. Thank you Skadi, Karen, Sara and Hoda for all the fun times we had while working, at the coffee machine, at the home-made diners and at the party's. Skadi, bedankt voor je onuitputtelijke behulpzaamheid en betrokkenheid. De zeil trip was fantastisch! Sara, it was much fun sitting across you, because your happiness cheered me up. You were the only one I could discuss football with. Karen, je hebt me het echte Limburgse leven laten zien. En dat met een goede dosis humor. Hoda, thank you for all the deep discussions we had about life. You are a great Farsi teacher!

I would like to thank all my other research colleagues from the big room, Ralph, Stefan, Daniela, Adriana, Georgi, Scott, Davide, Shane, and others for creating such a pleasant working atmosphere. Shane, thank you for keeping us healthy with your endless supply of fruits and for the discussions we had about movies, series, and other (controversial) topics. Ralph, bedankt voor al je hulp met de radiomics tool en je sterke verhalen.

De hele Maastricht fysica groep wil ik ook bedanken voor hun betrokkenheid. Michel, altijd lekker om even een bakkie te doen en de dag door te nemen. Fiere, de vele treinreizen tussen Eindhoven en Maastricht waren een stuk fijner als we samen in de trein zaten. Bedankt voor het vertrouwen dat ik alles met je kon bespreken en voor je vrolijkheid en enthousiasme. Wanneer gaan we weer een chocomel doen?

Ook wil ik graag de gehele medische staf van Maastricht bedanken voor hun inzet voor de prospectieve trials. Daarbij bedank ik ook alle patiënten die betrokken zijn geweest bij dit onderzoek. Zonder hen zou deze bijdrage aan de kennis over kanker niet tot stand zijn gekomen. Jeroen, bedankt voor je benaderbaarheid voor alle medische vragen die ik had. In ruil daarvoor kreeg je statistisch advies terug. Het was erg plezierig om met je samen te werken op de endeldarm kanker projecten. Ludy, de cervix en endometrium projecten waren interessant en uitdagend. Bedankt voor je heldere uitleg daarbij. Maaïke en Patricia, gezellig dat ik een beetje op de hoogte kon blijven van de AIOS kamer.

Alle laboranten die betrokken zijn geweest bij het behandelen en scannen van patiënten voor de betreffende studies: bedankt. Marlies, het was altijd een feest om even met de harde schijf naar de PET scanner te gaan als jij aan het werk was. Jorgen, thanks voor je werk voor de observatie studies en je input tijdens de abdomen meetings.

Daarnaast wil ik alle datamanagers bedanken voor al het werk dat ze voor de prospectieve studies hebben verricht. Ann, je was altijd erg behulpzaam om alle databases compleet te krijgen. Sylvia, bedankt dat je alle data uit de THUNDER trial hebt verzameld. Cary, het was fijn om met je samen te werken en expertise uit te wisselen over voorspellingsmodellen. Jouw kennis en zorgvuldigheid als onderzoeker hebben vele projecten vooruit geholpen.

Andre, bedankt voor je scherpe blik en nuttige input voor mijn onderzoek. Jouw visies over het delen van data en gedistribueerde model training hebben me zeker geïnspireerd. Johan, fijn dat je een vervolg geeft aan de endeldarmkanker onderzoeksprojecten.

Ik wil ook alle studenten (Joost, Maja, Veerle, Michel) bedanken die me op welke manier dan ook hebben geholpen bij het onderzoek. Veerle, naast dat je hard gewerkt hebt, was het vooral ook erg gezellig om je te begeleiden.

Francesco, thank you for our collaboration. I have learned a lot from you and I was happy to support you in your research at Maastricht. I hope to see you again in Italy some day. In bocca al lupo!

Ook wil ik alle medewerkers van Maastricht lab bedanken. De activiteiten die we samen hebben georganiseerd waren erg leuk. Ook voelde ik me altijd welkom om even koffie bij jullie te komen drinken.

Graag wil ik alle onderzoekers bedanken van zowel de afdeling Knowledge Engineering in Maastricht als het Tilburg center for Cognition and Communication in Tilburg. Jeroen en Giel, bedankt voor de gezelligheid in de eerste 9 maanden als mijn kamergenoten in Maastricht en daarna in Tilburg. Vooral de foute filmavonden, de sushi en de feestjes zijn me bijgebleven van deze mooie tijd.

In de zes ESTRO's waar ik bij geweest ben, heb ik vele boeiende mensen leren kennen. Maar er was altijd een kleine harde kern waar ik mee op de dansvloer stond. Bedankt Anette, Ellen, Lia, Michiel, Tezontl en al jullie andere collega's!

I would like to thank all the people in Harris' lab and the rest of the Oncology Department in Oxford for the last year while writing up the thesis. Francesca, thank you for giving me the opportunity to learn this new field and for your patience in the last year of my PhD. I admire the way you managed to set up such an exciting research group. Syed and Sheng, it is a pleasure to work among such good friends. Let's make some more trips in the UK now!

Dear housemates at Rewley Road (Philip, Allison, Sara and Sharon), I consider myself very lucky to have found a house with such lovely people. Thank you for everything.

Loes, bedankt voor al die jaren van steun en afleiding die je me hebt gegeven tijdens deze promotie. Het was een bijzonder mooie tijd in Eindhoven. Jan, Annie, Wout, Pleun en Erik, ook jullie bedankt voor alles. En niet te vergeten de familie Ruijs en de familie Roefs voor alle gezellige bijeenkomsten.

Mijn BMT studie vrienden (en hun vriendinnen), Huy, Rudy, Duy, Joost, Edwin, Wouter, Joris en Marcel, bedankt voor de ontspannen avonden en de activiteiten die we georganiseerd hebben zodat ik even niet aan mijn promotie hoefde te denken. Huy en Joost, thanks voor alle etentjes en feestjes in Maastricht. En ook fijn dat er altijd een slaapplekkie was. Duy, bedankt voor je behulpzaamheid en gastvrijheid in Eindhoven. Rudy, het was altijd erg gezellig om bij jullie langs te komen.

Wolter, Renske, Anke, Andries, ondanks dat ik jullie niet meer zo vaak zag als toen in Eindhoven, waren die keren dat we elkaar zagen weer als vanouds. Bedankt voor jullie betrokkenheid en humor.

De vrienden die ik echt al langer ken zijn gelukkig ook gebleven. Rob, Ilse, Niels en Jeannet, bedankt voor de gezellige etentjes, spellenavonden, festivals, kermissen, en andere leuke activiteiten die we in de laatste jaren hebben ondernomen.

All the (former) members of ESBV Tantalus that I met, thanks a lot for all the fun on and outside the basketball court. The Friday nights were epic and the IBT tournaments were splendid (and other tournaments too). It is remarkable how the love for a game can connect people to this extend.

Tot slot wil ik mijn familie bedanken voor hun interesse in mijn werk, de steun in de laatste jaren en de mooie momenten die we hebben mogen delen. Een aantal van deze mensen wil ik graag speciaal vermelden.

Ciel, Anne, Bart, Fieke, Bas en Bram, bedankt dat jullie mij zo hartelijk hebben ontvangen in jullie liefdevolle familie. Ik voelde me vanaf het begin meteen op m'n gemak. Het is iedere keer weer erg gezellig met z'n allen.

Petra en Kees, bedankt dat jullie als peetoom en peettante deze traditie nog steeds in stand hebben gehouden. De Brabantse gezelligheid op jullie feestjes kan ik zeker waarderen.

Tom en Joyce, ik vind het heel leuk om te zien dat jullie zo gelukkig zijn samen. Ik kijk ook al uit naar jullie bruiloft (en vereerd om getuige te zijn). Broer, bedankt voor je steun in al die jaren. We hebben een hoop lol gehad in onze fantasiewerelden. Ik had geen betere broer kunnen wensen!

Pap en mam, bedankt voor al jullie betrokkenheid, steun en interesse tijdens mijn promotie. Mam, bedankt dat je mij altijd hebt gestimuleerd om zo ver mogelijk door te leren. Ik weet hoe trots je bent op dit resultaat. Het is altijd heerlijk thuiskomen in Gestel waar ik helemaal mezelf kan zijn. Pap, bedankt dat je mij gestimuleerd hebt om kritisch en vanuit verschillende perspectieven na te denken over alles wat er in de wereld gebeurt. Ik bewonder de proactieve manier waarop je je leven inricht.

Lieve Sarah, bedankt voor al je steun, begrip, blijdschap, en liefde in het afgelopen jaar. Je maakt me altijd aan het lachen waardoor ik alle stress even vergeet. Ik heb je op de proef gesteld door naar Oxford te verhuizen, maar dat houdt onze zonnige toekomst samen niet tegen!



# Curriculum Vitae

## Curriculum Vitae

Ruud van Stiphout was born on August 31<sup>st</sup> 1982 in 's-Hertogenbosch, The Netherlands. He spent his childhood in Sint Michielsgestel and in this village he finished high school at Gymnasium Beekvliet in 2000. That same year he started Biomedical Engineering at the Eindhoven University of Technology. After obtaining the Bachelor's degree in 2003, he started the Master Biomedical Engineering in the Biomedical Imaging and Modeling group. His projects mainly involved modeling of cellular processes in time using systems biology approaches. Part of his program was the internship at the Semmelweis University in Budapest (Hungary), where he studied calcium dynamics in heart muscle cells. His final project was on calcium induced activation mechanisms of skeletal muscle cells. He finished his Masters in 2007 'with great appreciation'. In 2008 he started as a PhD student at both Maastricht clinic and the Department of Knowledge Engineering at Maastricht University. His research involved the development of prediction models for treatment outcome in rectal cancer, of which the results are presented in this thesis. From February 2013 he is working as a post-doctoral researcher on applied computational genomics projects at the University of Oxford in the United Kingdom.

## List of publications

## List of publications

### *Journal publications*

- **van Stiphout RGPM**, Valentini V, Buijsen J, Lammering G, Meldolesi E, Gambacorta MA, Dekker A, Lambin P. Nomogram predicting response after chemoradiotherapy in rectal cancer using sequential PETCT imaging: a prospective study with external validation (submitted)
- **van Stiphout RGPM\***, Valentini V\*, Lammering G, Gambacorta MA, Barba MC, Bebenek M, Bonnetain F, Bosset JF, Bujko K, Cionini L, Gerard JP, Rödel C, Sainato A, Sauer R, Minsky B, Collette L, Lambin P. Two-year disease-free and pathological complete response as clinical intermediate endpoint using pooled data of randomized trials for locally advanced rectal cancer (submitted)
- **van Stiphout RGPM\***, Buijsen J\*, Menheere PCA, Lammering G, Lambin P. Blood biomarkers are helpful in the prediction of response to chemoradiation in rectal cancer: a prospective, hypothesis driven study on patients with locally advanced rectal cancer (submitted)
- Roelofs E, Dekker A, Meldolesi E, **van Stiphout RGPM**, Valentini V, Lambin P. International data-sharing for radiotherapy research: An open-source based infrastructure for multicentric clinical data mining. *Radiother Oncol.* 2013 [Epub ahead of print]
- Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijsen J, Zegers CM, Carvalho S, Leijenaar RT, Nalbantov G, Oberije C, Scott Marshall M, Hoebbers F, Troost EG, **van Stiphout RGPM**, van Elmpt W, van der Weijden T, Boersma L, Valentini V, Dekker A. 'Rapid Learning health care in oncology' - an approach towards decision support systems enabling customised radiotherapy'. *Radiother Oncol.* 2013 ;109(1):159-64.
- **van Stiphout RGPM\***, Lambin P\*, Starmans MH, Rios-Velazquez E, Nalbantov G, Aerts HJ, Roelofs E, van Elmpt W, Boutros PC, Granone P, Valentini V, Begg AC, De Ruyscher D, Dekker A. Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nature Rev Clin Oncol.* 2013;10(1):27-40

- Hatt M, **van Stiphout RGPM**, le Pogam A, Lammering G, Visvikis D, Lambin P. Early prediction of pathological response in locally advanced rectal cancer based on sequential (18)F-FDG PET. *Acta Oncol.* 2013;52(3):619-26.
- Buijsen J, van den Bogaard J, van der Weide H, Engelsman S, **van Stiphout RGPM**, Janssen M, Beets G, Beets-Tan R, Lambin P, Lammering G. FDG-PET-CT reduces the interobserver variability in rectal tumor delineation. *Radiother Oncol.* 2012;102(3):371-6.
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, **van Stiphout RGPM**, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48(4):441-6.
- Janssen MH, Öllers MC, **van Stiphout RGPM**, Riedl RG, van den Bogaard J, Buijsen J, Lambin P, Lammering G. PET-based treatment response evaluation in rectal cancer: prediction and validation. *Int J Radiat Oncol Biol Phys.* 2012;82(2):871-6.
- **van Stiphout RGPM\***, Valentini V\*, Lammering G, Gambacorta MA, Barba MC, Bebenek M, Bonnetain F, Bosset JF, Bujko K, Cionini L, Gerard JP, Rödel C, Sainato A, Sauer R, Minsky BD, Collette L, Lambin P. Nomograms for predicting local recurrence, distant metastases and overall survival for locally advanced rectal cancer patients based on European randomized clinical trials. *J Clin Oncol.* 2011; 29(23):3163-72
- **van Stiphout RGPM**, Lammering G, Buijsen J, Janssen MH, Gambacorta MA, Slagmolen P, Lambrecht M, Rubello D, Gava M, Giordano A, Postma EO, Haustermans K, Capirci C, Valentini V, Lambin P. Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential PET-CT imaging. *Radiother Oncol* 2011;98:126-133.
- Lambin P, Petit SF, Aerts HJ, van Elmpt WJ, Oberije CJ, Starmans MH, **van Stiphout RGPM**, van Dongen GA, Muylle K, Flamen P, Dekker AL, De Ruyscher D. 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. *Radiother Oncol.* 2010;96(2):145-52.

- Janssen MH, Ollers MC, **van Stiphout RGPM**, Riedl RG, van den Bogaard J, Buijsen J, Lambin P, Lammering G. Blood glucose level normalization and accurate timing improves the accuracy of PET-based treatment response predictions in rectal cancer. *Radiother Oncol.* 2010;95(2):203-8.
- Janssen MH, Ollers MC, **van Stiphout RGPM**, Buijsen J, van den Bogaard J, de Ruyscher D, Lambin P, Lammering G. Evaluation of early metabolic responses in rectal cancer during combined radiochemotherapy or radiotherapy alone: sequential FDG-PET-CT findings. *Radiother Oncol.* 2010;94(2):151-5.
- Janssen MH, Ollers MC, Riedl RG, van den Bogaard J, Buijsen J, **van Stiphout RGPM**, Aerts HJ, Lambin P, Lammering G. Accurate prediction of pathological rectal tumor response after two weeks of preoperative radiochemotherapy using (18)F-fluorodeoxyglucose-positron emission tomography-computed tomography imaging. *Int J Radiat Oncol Biol Phys.* 2010;77(2):392-9.
- **van Stiphout RGPM**, van Riel NA, Verhoog PJ, Hilbers PA, Nicolay K, Jeneson JA. Computational model of excitable cell indicates ATP free energy dynamics in response to calcium oscillations are undamped by cytosolic ATP buffers. *Syst Biol (Stevenage).* 2006;153(5):405-8.
- Ligeti L, Szenczi O, Prestia CM, Szabó C, Horváth K, Marcsek ZL, **van Stiphout RGPM**, van Riel NA, Op den Buijs J, Van der Vusse GJ, Ivanics T. Altered calcium handling is an early sign of streptozotocin-induced diabetic cardiomyopathy. *Int J Mol Med.* 2006 Jun;17(6):1035-43.

### ***Book chapters***

- **van Stiphout RGPM**, Roelofs E, Dekker A, Lambin P. How Should Data Be Shared and Rapid Learning Health Care Promoted? (Multidisciplinary Management of Rectal Cancer – Questions and Answers, Springer, 2012)
- Vincenzo Valentini V, Francesco Cellini F, Maria Cristina Barba MC, **van Stiphout RGPM**. Do Different Populations of Rectal Cancer Exist? (Multidisciplinary Management of Rectal Cancer – Questions and Answers, Springer, 2012)

\* equal contribution to the work