Statistical methods to analyze treatment strategies from randomized and non-randomized studies in oncology

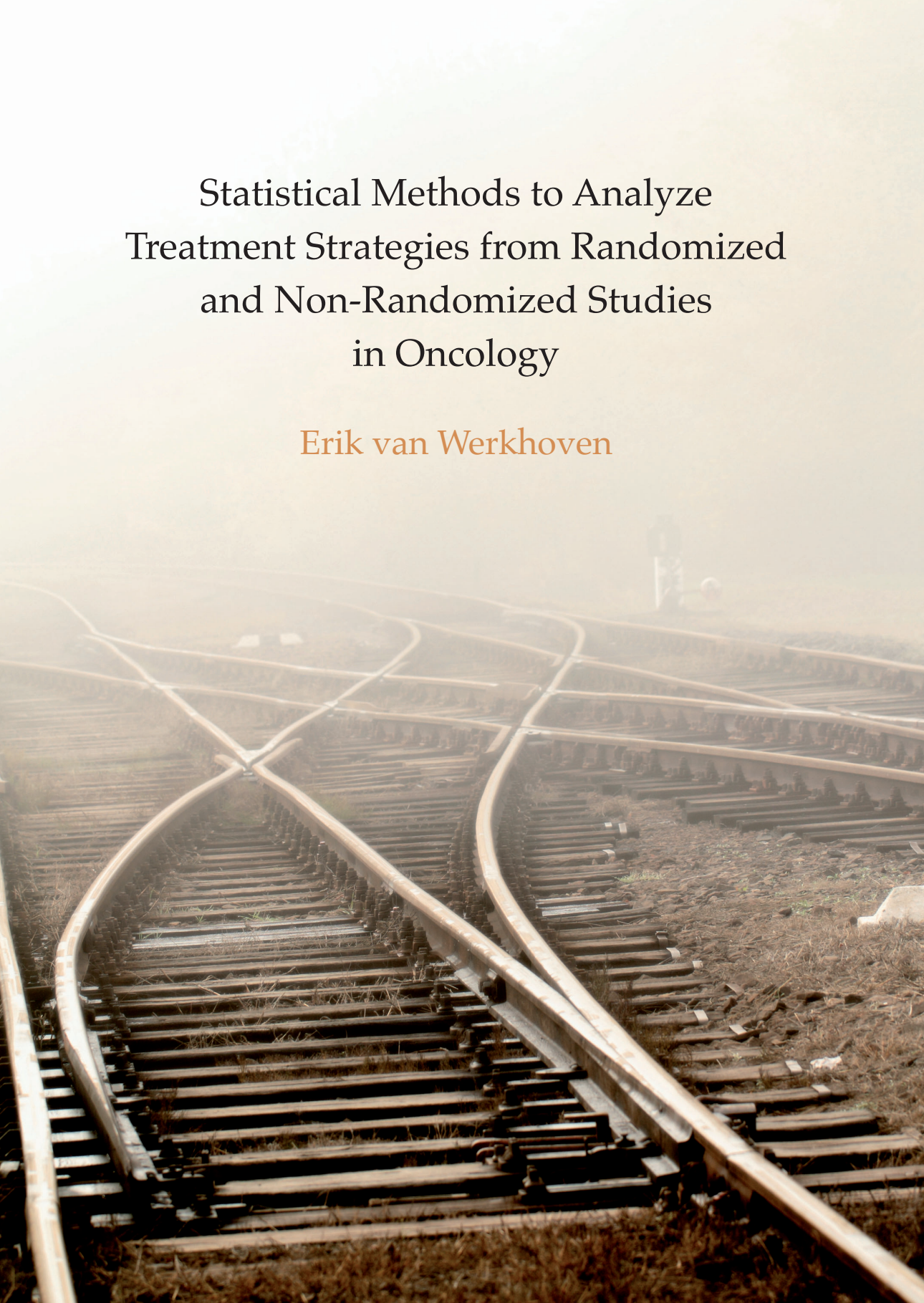
van Werkhoven, E.

Publication date
2023

Document Version
Final published version

Statistical Methods to Analyze Treatment Strategies from Randomized and Non-Randomized Studies in Oncology

Erik van Werkhoven



Statistical Methods to Analyze Treatment Strategies from Randomized and Non-Randomized Studies in Oncology

Erik van Werkhoven

The publication of this thesis was financially supported by the Academic Medical Center, University of Amsterdam.

Printed by: Optima Grafische Communicatie, Rotterdam
Cover photo: Getty Images
Cover design: Erwin Timmerman / Optima Grafische Communicatie, Rotterdam

Statistical Methods to Analyze Treatment Strategies from Randomized and Non-Randomized Studies in Oncology

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P. P. C. C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 20 september 2023, te 10.00 uur

door Erik Daniël van Werkhoven
geboren te Woerden

Promotiecommissie

Promotores:	prof. dr. P. M. M. Bossuyt	AMC-UvA
	prof. dr. A. H. Zwinderman	AMC-UvA
Overige leden:	prof. dr. L. J. A. Stalpers	AMC-UvA
	dr. M. G. H. van Oijen	AMC-UvA
	prof. dr. J. J. Cornelissen	Erasmus Universiteit Rotterdam
	prof. dr. S. le Cessie	Universiteit Leiden
	prof. dr. K. C. B. Roes	Radboud Universiteit
	prof. dr. M. G. H. Besselink	AMC-UvA

Faculteit der Geneeskunde

Contents

List of Abbreviations

CI	Confidence interval
df	Degrees of freedom
HR	Hazard ratio
IQR	Interquartile range
NA	Not available / missing value
w/	With
w/i	Within
w/o	Without

CHAPTER 1

Introduction

Shared decision making is an important part of personalized care. It is a process in which "clinicians and patients work together to select tests, treatments, management or support packages, based on clinical evidence and the patient's informed preferences" (Coulter & Collins, 2011).

To help clinicians choose a treatment together with the patient, taking the patients' individual characteristics into account, this thesis explores statistical methods to analyze the effects of treatment strategies from both randomized and non-randomized studies in oncology. Below we will explain why the difference between these two types of studies is consequential, and why statistical methods to estimate treatment effects from non-randomized studies have gained importance.

The accuracy of estimates of treatment effects is threatened by two types of error: random error and systematic error. Random error, or noise, refers to the error that is caused by the fact that values that are estimated from samples instead of the total population, will vary from one sample to another. Random error can be reduced by increasing the sample size.

Unlike random error, systematic error, or bias, cannot be reduced by increasing the sample size. There are many possible causes of bias—the online Catalogue of Bias currently contains 62 entries (The Catalogue of Bias Collaboration, 2022). One form of bias is treatment-selection bias. This is the bias that occurs if treatment groups are compared, but the treatment selection is influenced by patient's characteristics. Random treatment allocation can eliminate this bias, and is perhaps the only reliable method to do so (Bosco et al., 2010).

Randomized clinical trials are therefore considered to provide the highest level of

scientific evidence in medicine. The randomization process allows isolation of treatment effects, because if the sample size is large enough, it reduces bias by balancing the distribution of patients' characteristics across the treatment groups. This holds for known predictors (confounders), and even for unobserved or unknown factors (Piantadosi, 2017, p. 76).

Although randomized clinical trials are the preferred source of scientific evidence, it is not always possible to conduct them. There are many possible reasons. For example in pediatric studies, parents may be reluctant to enroll their child in a trial if he or she might receive a placebo (Augustine et al., 2013). Another example are new treatments that are tested for diseases for which giving placebo treatment is considered unethical. This may be the case if there is an 'unmet medical need'. The European Commission has defined this as "a condition for which there exists no satisfactory method of diagnosis, prevention or treatment [...] or, even if such a method exists, in relation to which the medicinal product concerned will be of major therapeutic advantage to those affected" (EC 507/2006). The European Medicines Agency (EMA) can give conditional marketing authorization in such cases.

A more practical reason is that randomized trials require larger sample sizes than single-arm trials. Larger sample sizes are a concern from an ethical perspective, because patients who participate in a trial take the risk of potentially getting a suboptimal treatment compared to standard of care. Furthermore, larger sample sizes mean that these trials cost more than single-arm studies, and that take longer time before they can be analyzed. Especially in fields of research that study rare diseases, it is often argued that it is impossible to conduct randomized trials for this reason (Lasch et al., 2017).

In absence of any control data, it is only possible to evaluate the outcome of a single-arm study if the natural course of the disease is very well known (CHMP, 2006). In practice, uncertainty at the design stage about the choice of the null hypothesis may increase the risk of a type I error (a false positive trial) if the null effect is underestimated, and type II error (a false negative trial) if the null effect is overestimated (Foster et al., 2020).

The difficulty of specification of the null hypothesis will be even greater if the outcome of the trial is based on a time-to-event endpoint, like progression-free survival (PFS) or overall survival (OS) (Foster et al., 2020), because of the variability in the natural history of many diseases. The FDA even recommends to choose objective response and complete response as dichotomous endpoints for single-arm studies in oncology, instead of PFS or OS (FDA, 2018).

One way to get around the problem that the natural history of the disease with standard treatment must be known, if no randomized trial can be conducted, is to do a single-arm study and strengthen it with observational data in the role of

a control group. Aside from knowledge of the course of the disease, Piantadosi (2017) lists four more requirements for the reliability of inferences about treatment effects from non-experimental data:

- there must be patients who get the treatment of interest in the same way as we intend to study it;
- the relevant outcome variables must be available for these patients;
- the effect of the treatment or intervention must be large relative to random error and bias (for example, selection bias or indication bias);
- evidence of efficacy must be consistent with other biological knowledge.

These are prerequisites for any analysis that involves a comparison of non-randomized treatments, whether observational data are used instead of a control group, or both treatment groups are obtained from observational data.

In this thesis, data from different types of sources are analyzed: data from randomized trials, from registries, and data from single-arm studies augmented with registry data. For the latter two, the possibilities are explored to obtain clinical evidence from these types of studies using various statistical methods to reduce treatment-selection bias. The focus will be on three oncological diseases: breast cancer, lymphoma, and ovarian cancer.

The thesis starts with an analysis of a randomized trial for early-stage breast cancer patients, for whom the standard of care is breast-conserving therapy. This therapy consists of breast-conserving surgery, followed by whole-breast irradiation. It has been demonstrated in the EORTC 22881-10882 Boost/No Boost randomized trial that the addition of an extra radiation 'boost' dose on the original tumor bed decreased the risk of local tumor recurrence. In **chapter 2** a prediction model is developed to predict the probability of a local recurrence at 10 years, using the Boost/No Boost trial, including reviewed pathology data. The aim is to offer predictions of the chance of a local recurrence in a scenario with and without boost to individual patients, using their particular demographic and tumor characteristics. In **chapter 3** an additional analysis of the Boost/No Boost trial is performed after collection of longer follow-up data.

The successor of the Boost/No Boost trial is the Young Boost trial, which investigates a further increase of the boost dose in patients of 50 years and younger. Cosmetic outcome is collected in this trial as a patient-reported outcome from questionnaires. In addition, it is scored by the treating physicians and analyzed by a computer program, called BCCT.core, which uses digital photographs to generate a score. In **chapter 4** the agreement between these three scores is investigated. Associations of baseline factors with cosmetic results and incidence of fibrosis are analyzed in **chapter 5**.

Chapter 6 presents an analysis to investigate trends over time in treatment outcome for early-stage breast cancer. A non-experimental cohort is used that consists of pa-

tients who were treated with breast-conserving therapy at the Netherlands Cancer Institute between 1980 and 2008.

In **chapter 7**, several methods are explored to estimate a treatment effect from a single-arm phase II study. The study evaluated the addition of lenalidomide to the standard R-CHOP treatment for patients with diffuse large B-cell lymphoma that harbors a *MYC*-rearrangement.

In **chapter 8** an existing treatment-selection rule is externally validated in an observational cohort. The rule was been developed to aid in the choice between primary surgery and neoadjuvant chemotherapy for treatment of patients with advanced epithelial ovarian cancer.

Some ideas about possible directions for future research are presented in **chapter 9**.

CHAPTER 2

Nomogram to Predict Ipsilateral Breast Relapse Based on Pathology Review From the EORTC 22881–10882 Boost Versus No Boost Trial

This chapter was published as:
Erik van Werkhoven, Guus Hart, Harm van Tinteren, Paula Elkhuizen, Laurence Collette, Philip Poortmans, Harry Bartelink
Nomogram to predict ipsilateral breast relapse based on pathology review from the EORTC 22881–10882 boost versus no boost trial
Radiotherapy and Oncology 100(1):101–107, 2011.

Abstract

Background and purpose The EORTC 22881-10882 trial showed that for patients treated with breast conserving therapy (BCT), a 16 Gy boost dose significantly improved local control, but increased the risk of breast fibrosis. A model to estimate the risk of ipsilateral breast relapse (IBR) already exists, but now a model has been developed which takes boost treatment into account and is based on centrally reviewed pathology.

Materials and methods A Cox model was developed based on central pathology review data and clinical data of 1603 patients from the EORTC 22881-10882 trial with a median follow-up of 11.5 years. From a predefined set of variables, predictors with a maximal effect on 10-year IBR rate $> 4\%$ were retained in the model. Bootstrap re-sampling was used to assess model calibration and discrimination. The results are presented in the form of a nomogram.

Results Apart from young age and no boost, presence of DCIS adjacent to the invasive tumor was associated with increased risk of IBR (HR 1.96, $p = 0.001$). Patients with high grade invasive tumors were younger than patients with low/intermediate grade ($p < 0.0001$). The nomogram includes histologic grade, DCIS, tumor diameter, age, tamoxifen, chemotherapy, and boost with a concordance probability estimate of 0.68.

Conclusions The nomogram for predicting IBR 10 years after BCT includes seven factors, with young age, presence of DCIS and boost treatment as the most dominant factors. The nomogram estimates IBR and confirms the importance of a boost dose. Combined with a model to predict fibrosis published previously, the nomogram presented here may assist in decision making for individual patients.

2.1 Introduction

Nowadays, breast-conserving therapy (BCT) for most early breast cancer patients. A meta-analysis with long term follow-up demonstrated equal survival rates between radical mastectomy and BCT for early breast cancer patients, although a slightly higher ipsilateral breast relapse (IBR) rate was observed in the BCT patients (EBCTCG, 2005). The EORTC 22881-10882 trial with 5318 conservatively treated early breast cancer patients showed that a 16 Gy boost dose significantly reduced IBR, although at the cost of an increased risk of breast fibrosis (Bartelink et al., 2001, 2007) Young age is one of the most important factors related to the local recurrence rate as was seen in the EORTC trial, as well as in many other studies (Antonini et al., 2007; Bollet et al., 2007; Beadle et al., 2011; Elkhuizen et al., 2000). Involvement of margins was also reported to be related to a higher local recurrence rate (Poortmans et al., 2009; Kreike et al., 2009). However, the negative impact of close or involved margins was not significant in other publications (Morrow, 2008). For example, margin status was not statistically significant in the multivariable analysis of local recurrence in the EORTC boost versus no boost trial (Jones et al., 2009). Three factors that remained significant were young age, histologic grade, and boost dose. Fortunately, a higher radiation dose reduces the local recurrence rate, an effect which appeared to be independent on the hazard ratio scale in all age groups, with the largest absolute benefit of a boost dose in young women.

Nomograms have been developed to estimate individual patient treatment outcome. Tufts investigators constructed a nomogram for patients with early breast cancer which was based on randomized clinical trials and institutional studies. The web-based version was named "IBTR!" (Sanghani et al., 2007). Features of this nomogram are age, tumor size and grade, margin status, lymphovascular invasion, and systemic adjuvant treatment. This nomogram has recently been updated in IBTR! version 2 (Sanghani et al., 2010). In this nomogram, the 10 year risk of IBR was estimated, with and without radiotherapy of the whole breast. However, their model does not include information about an extra boost dose. Although a boost dose further reduces the IBR rate, it implies the disadvantage of carrying the risk of increased fibrosis and decreased cosmetic outcome. To estimate the risk of fibrosis in BCT, we recently developed a nomogram based on the EORTC boost versus no boost trial (Collette et al., 2008). Major factors influencing fibrosis and cosmetic results are the radiation dose, the treatment technique for the boost dose and adjuvant chemotherapy.

The purpose of the present analysis is to develop a nomogram to estimate the 10-year IBR probability by using pathology features from the review pathology (such as the presence of DCIS adjacent to the invasive tumor), treatment factors (systemic adjuvant treatment with hormones, chemotherapy, and boost dose), and age. We propose to use both nomograms jointly for risk prediction of IBR and fibrosis in the

clinic, for guiding decisions for individual patients.

2.2 Materials and Methods

Study Population

In the EORTC 22881-10882 (boost versus no boost) trial 5318 patients were randomized between no boost and a 16 Gy boost dose (Bartelink et al., 2007). From the early years of the accrual period (1989–1996), pathology slides from 1616 patients (which represent 30% of the whole population) were collected and reviewed by a single pathologist (Hans Peterse). Thirteen patients who were ineligible, whose tumor was incompletely resected, or with too much data missing were removed, leaving 1603 patients in this analysis.

Prognostic Variables

Variables were selected in two steps. First the following centrally reviewed pathology variables were selected based on clinical expertise and supported by univariable analysis: largest diameter of dominant lesion, histologic grade of invasive tumor according to the Elston/Ellis modification of the Bloom-Richardson system, and type of in situ tumor adjacent to dominant lesion, which was reclassified as presence of DCIS or not. Additionally, the following non-centrally reviewed variables were selected: age at randomisation, presence of positive nodes, chemotherapy, tamoxifen, and boost treatment.

After a model was developed using these pre-selected variables, a backward variable selection procedure was used in which only clinically relevant variables (i.e. with a maximum absolute effect size of 4% or more) were kept. The effect of each variable was estimated at 10 years by setting all other variables at the worst predicted level (for factors) or the worst percentile (for continuous variables). Histologic grade was coded by dummy variables, so that the model selection mechanism decided which factor levels to combine.

Statistical Methods

Baseline patient characteristics were compared and screened for collinearity using χ^2 and Wilcoxon's tests. Cluster analysis for missing values revealed no patterns.

Time was calculated from randomization to IBR as first event, without simultaneous regional or distant recurrence occurring within 4 months after IBR diagnosis. Other patients were censored at other breast cancer events, death or last follow-up. Cox models were used for univariable and multivariable analyses. Model assumptions were assessed by inspecting martingale and Schoenfeld's residuals. A restricted cubic spline with four knots was used for age, but it was also analyzed as

a binary variable with cut-off at 50 years for exploratory purposes (Harrell, 2001). A simultaneous test for pairwise interactions between boost treatment, age and the other candidate variables was performed.

Missing values were imputed to the median or the mode. A sensitivity analysis using a model fit on completely observed patient data only and an additional model on a dataset where missing values of DCIS were imputed to the least frequent observation ('No DCIS') showed no major discrepancies. The final model was presented in the form of a nomogram.

Model Validation

Internal validation of the model was performed using the bootstrap procedure as advised by Harrell et al.: (1) draw a random sample from the original data with replacement and of the same size as the original data, (2) develop a model on this bootstrap sample using the variable selection mechanism, (3) calculate the difference in model performance between the bootstrap and the original dataset, (4) repeat steps (1–3) 500 times and average the differences found (Harrell et al., 1996). This average is an estimate of the optimism and can be subtracted from the original model performance. An additional analysis was performed using a modification of this procedure where in step (3) the model from the bootstrap sample was tested using only the data of patients not selected in that particular bootstrap sample (instead of the complete original dataset).

The Gonen and Heller concordance probability estimate (CPE) was calculated to assess model discrimination. For two patients, one of whom had an IBR and the other did not by a certain time, the CPE estimates the probability that the model will give higher risk to the one patient instead of the other. A model with perfect discrimination would have a concordance probability of 1, whereas a value of 0.5 indicates that a coin toss would provide information as accurate as that given by the model (Gönen & Heller, 2005). A calibration plot was drawn showing predicted 10-year recurrence free probabilities against observed Kaplan-Meier estimates, grouped into intervals containing 200 subjects on average. Analyses were performed using SAS software (version 9.2, SAS Institute Inc., Cary, NC) and R (version 2.11.1, http://www.R-project.org/) with packages rms and CPE (Harrell, 2001; R Core Team, 2010; Mo et al., 2010).

2.3 Results

The median follow-up for the patients in the subset of 1603 patients was 11.5 years. There were 120 ipsilateral breast relapses (IBR) as first event. Table 2.1 shows first events by treatment arm. The median age was 54 years (range 27–76, interquartile range (IQR) 16 years), and the median tumor diameter 15 mm (range 0–50, IQR

9 mm). One tumor had a recorded diameter of zero. Univariable analysis showed boost treatment and tamoxifen administration were significantly associated with a lower risk of local recurrence; while young age, high histologic grade, and presence of DCIS adjacent to the dominant lesion or DCIS involved in the margin were related with a higher risk (Table 2.2). Collinearity between age and histologic grade was found: patients with high grade invasive tumors were younger (median 50 years) than patients with low/intermediate grade (median 55 years, $p < 0.0001$, figure 2.1).

Table 2.1: First events by treatment arm. The event of interest was IBR. Patients with other events or with simultaneous regional or distant recurrence occurring within 4 months after IBR diagnosis were censored.

	No Boost 795	**16 Gy Boost** 808	**Total** 1603
No event	483	529	1012
Ipsilateral breast relapse (IBR)	78	42	120
Regional recurrence	17	19	36
New contralateral tumor	35	30	65
Other new primary tumor	51	42	93
Distant metastasis	93	111	104
Death	29	29	58
IBR simultaneous with other recurrence (w/i 4 months)	9	6	15

Multivariable analysis showed histologic grade as a significant prognostic variable when age was modeled as binary variable, with the hazard ratio (HR) of high grade versus low/intermediate grade 1.64 (95% CI 1.09–2.46, $p = 0.02$). However, when age was modeled continuously the HR was 1.42 (0.94–2.14, $p = 0.10$). A global test for assessing interactions was decisive ($p > 0.4$) and therefore there was no need to repeat this test in each bootstrap sample during model validation. An interaction between boost treatment and DCIS ($p = 0.04$) was not included in the model because of the global test result.

Table 2.2: Univariable Cox analysis of IBR as first event

	Events	Patients	HR	95% CI	p-value
Tamoxifen					0.0011
No	108	1234	1		
Yes	12	369	0.39	0.21–0.70	

(Continued on next page)

Table 2.2 – continued from previous page

	Events	Patients	HR	95% CI	p-value
Chemotherapy					0.77
No	100	1351	1		
Yes	20	252	1.08	0.67–1.74	
Randomised treatment					0.0007
No Boost	78	795	1		
16 Gy Boost	42	808	0.53	0.36–0.77	
Histology invasive tumor		(34 missing)			0.41
Ductal	87	1116	1		
Lobular ca.	6	109	0.70	0.31–1.60	
Mixed pattern	8	163	0.60	0.29–1.23	
Other	12	181	0.80	0.44–1.46	
Histologic grade		(74 missing)			0.0052
Low	48	778	1		
Intermediate	25	392	1.15	0.71–1.86	
High	37	359	1.98	1.29–3.03	
Histologic grade (grouped)		(74 missing)			0.0014
Low / Intermediate	73	1170	1		
High	37	359	1.89	1.27–2.81	
Vascular invasion		(43 missing)			0.13
None	88	1179	1		
Present	20	222	1.30	0.80–2.11	
Doubtful	6	159	0.51	0.23–1.17	
MAI		(23 missing)			0.12
≤ 9	76	1149	1		
10–19	18	210	1.41	0.84–2.35	
≥ 20	20	221	1.58	0.97–2.59	
DCIS		(38 missing)			0.003
No	32	660	1		
Yes	81	905	1.84	1.22–2.77	
Margin of DCIS		(102 missing)			0.74
Free	42	472	1		
Close	17	215	0.96	0.54–1.68	
Involved	13	116	1.25	0.67–2.32	
Histologic grade of DCIS		(3 missing)			0.16
High	31	281	1		
Intermediate	35	401	0.76	0.47–1.24	
Low	15	220	0.56	0.30–1.03	
Estrogen					0.31
Negative	30	334	1		
Positive	59	845	0.71	0.46–1.11	
Unknown	31	424	0.76	0.46–1.26	

(Continued on next page)

Table 2.2 – continued from previous page

	Events	Patients	HR	95% CI	p-value
Progesteron					0.89
Negative	29	378	1		
Positive	50	680	0.90	0.57–1.42	
Unknown	41	545	0.95	0.59–1.52	
Nodal status	(11 missing)				0.41
Node-negative	98	1236	1		
Node-positive	22	356	0.82	0.52–1.31	
Diameter	(23 missing)				0.16
(per 10 mm)	114	1580	1.20	0.93–1.54	
Margin of invasive tumor	(120 missing)				0.24
Close	19	304	1		
Free	89	1130	1.24	0.76–2.04	
Involved	1	49	0.31	0.04–2.31	
Age (dichotomous)					<0.0001
≤ 50	73	622	1		
≥ 51	47	981	0.398	0.28–0.57	

MAI: Mitotic activity index

Table 2.3 shows results of the final multivariable model, which included boost treatment (hazard ratio (HR) 2.02; 95% CI 1.39–2.95), presence of DCIS (HR 1.96, 1.30–2.94), tamoxifen administration (HR 1.70, 0.91–3.19 if not given), chemotherapy (HR 1.47, 0.89–2.43 if not given), high versus low/intermediate histologic grade (HR 1.21, 0.78–1.87), and tumor diameter (HR 1.13 per 10 mm, 0.86–1.47). The effect of age, adjusted for the other variables, is shown in figure 2.2.

The model is presented in the form of a nomogram in figure 2.3. Tumor diameter was assigned 0.39 points per mm (with 0 points for a hypothetical diameter of 0 mm), an age of 25 years 100 points (with 0 points for 80 years), no tamoxifen administration 17 points, no chemotherapy 13 points, no boost treatment 23 points, presence of DCIS 22 points, and high histologic grade six points.

The model was internally validated using bootstrap re-sampling. The apparent concordance probability estimate (CPE) was 0.685 and its bias-corrected value was 0.684 with the method advised by Harrell and 0.683 with the modified bootstrap, which tests every bootstrap-constructed model using only data of patients not in that particular bootstrap sample (Harrell et al., 1996). Figure 2.4 shows a calibration plot showing predicted 10-year recurrence free probabilities against observed Kaplan-Meier estimates, grouped into intervals containing 200 subjects on average.

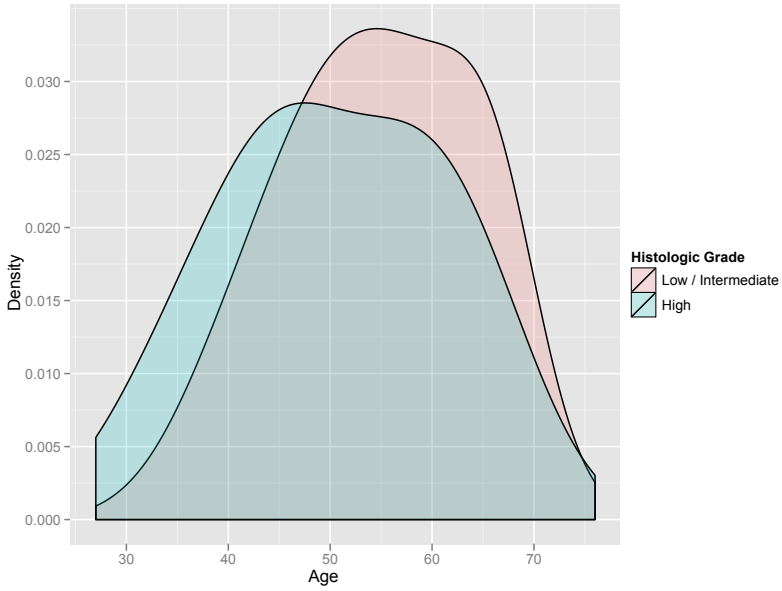


Figure 2.1: Kernel density estimates of distribution of age for low/intermediate histological grade (median 50 years) and high grade (median 55 years, $p < 0.0001$)

2.4 Discussion

The nomogram based on the Cox model reveals that young age, presence of DCIS and a boost dose of 16 Gy are the most important factors influencing the IBR rate after BCT for invasive stages I and II breast cancer. Other factors contributing to this nomogram are systemic treatment with chemotherapy or hormonal therapy, tumor diameter and high grade malignancy. This nomogram may guide the clinician to estimate the risk of ipsilateral breast relapse (IBR) in patients treated with BCT. Choices can be made for boost dose radiation or not, or if the risk of IBR remains high even for mastectomy. The previously developed nomogram can be used to estimate the risk of fibrosis after a boost dose following whole breast irradiation (Collette et al., 2008). In this way, by comparing the costs (fibrosis) and the benefits (local control), more informed decisions can be made together with the patient.

The most dominant factor contributing to the nomogram is age, a well known risk factor for IBR. The impact of young age on breast recurrences was shown in many other publications (Antonini et al., 2007; Bollet et al., 2007; Beadle et al., 2011). With the continuous variable for age we confirmed the collinearity between age and histological grade found in a previous analysis of this dataset, where a binary variable

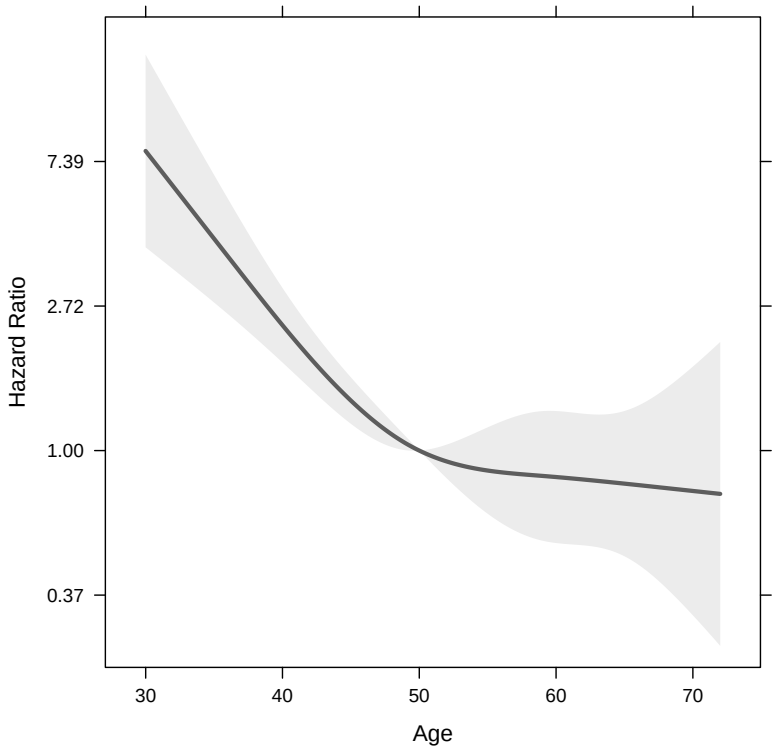


Figure 2.2: Spline curve for the effect of age in the multivariable model. The vertical axis has been graduated according to the log scale. The relative hazard has been set to 1 at the reference age of 50 years. The point-wise 95% confidence interval is shown by a shaded area

was used (Jones et al., 2009). Younger patients more often had a high grade malignancy. Exploration of grade using various modeling options for age showed that the significance of grade as predictor could vary. When age was considered as a binary indicator with a cut-off point at 50 years, the hazard ratio (HR) of high grade versus low/intermediate grade was 1.64 (95% CI 1.09–2.46, $p = 0.02$), and when it was considered continuous, the HR was 1.42 (0.94–2.14, $p = 0.10$). Therefore, in the former model, grade might have been included in the model to correct for the suboptimal modeling of the effect of age reduced to a binary. Another difference with the former model is that now the presence of DCIS was coded as a binary indicator, ignoring DCIS margin and grade.

The selection of model variables was separated into two steps. First, known clini-

Table 2.3: Final multivariable model

	HR	95% CI	p-value
Diameter			
(per 10 mm)	1.13	0.86–1.47	0.39
Age			
(per year)	(see figure 2.2)		< 0.0001
Tamoxifen			
Yes	1		
No	1.70	0.91–3.19	0.10
Chemotherapy			
Yes	1		
No	1.47	0.89–2.43	0.13
Randomised treatment			
16 Gy boost	1		
No boost	2.02	1.39–2.95	0.0003
DCIS			
No	1		
Yes	1.96	1.30–2.94	0.001
Histologic grade			
Low/intermediate	1		
High	1.21	0.78–1.87	0.39

cally important variables were selected based on literature findings, since both clinical and statistical significance are important to consider (Iasonos et al., 2008). In the second step variable selection was based on the maximal predicted effect on the absolute IBR rate at 10 years in a multivariable model. Effects smaller than 4% were left out of the nomogram, to make the model more parsimonious. This criterion is less stringent than a requirement for statistical significance, thus some factors were kept in the final model that did not reach the conventional level of statistical significance (diameter, tamoxifen, chemotherapy, and histologic grade). This seems reasonable, since these variables are well known to be prognostic. The backward stepwise selection method used here carries some limitations, since variables might be selected by chance alone. This is why validation by bootstrap re-sampling was used. Indeed, repeating the second step of the variable selection in every bootstrap sample ensured correct validation of model discrimination and calibration and enables insight in the effect of chance on the selection procedure. Age and boost treatment, for example, were selected in all 500 bootstrap replications, while diameter was chosen in only 331 out of 500 repetitions and grade (any of the two dummy variables) was selected 428 times. Collinearity between age and grade may make the estimates of the individual coefficients unstable, but reliable predictions

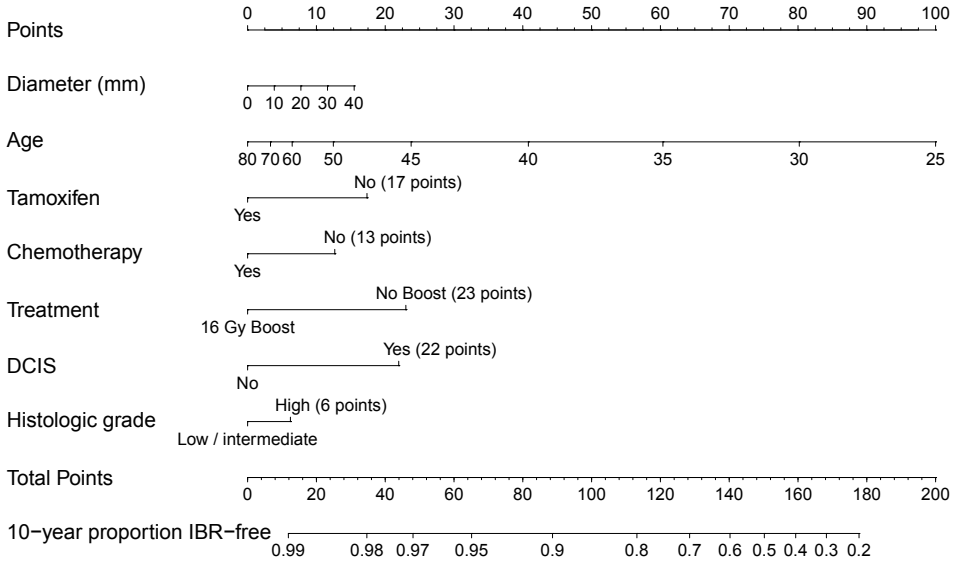


Figure 2.3: Nomogram predicting the 10-year IBR-free probability. First read off points for each variable along the points scale. Then add together the points and draw a vertical line from the total points scale to the 10 year IBR-free scale

can still be obtained (Steyerberg, 2009). Internal validation of the model showed only a small decline in CPE after bootstrap correction (from 0.685 to 0.683). With a CPE of around 0.68, the model is far from perfect, but, comparable to other cancer prediction models that achieved similar accuracy. A recently published model predicting IBR for patients with pure DCIS was reported to have a CPE of 0.69 (0.67 after correction) (Rudloff et al., 2010). Calibration estimates were also stable after correction (figure 2.4), although for the leftmost octile of patients (with lowest predicted recurrence-free rate), the predictions seem further off the diagonal compared to the other groups.

The variables selected for the model presented here differ from the ones used for the revised IBTR! 2.0 model by Sanghani et al. (2007). Table 2.4 presents an overview of the differences. Vascular invasion was not entered into the model because the IBR rate of patients whose vascular invasion was 'doubtful' (159 patients) seemed not to be estimated accurately in the univariable analysis (table 2.2). Margin involvement was omitted considering the high number of missing values (120 patients). It has to be mentioned here that all patients used for this analysis had a complete excision of the invasive tumor according to the local pathologist. Some other studies investigating the significance of microscopic resection margins also found only lim-

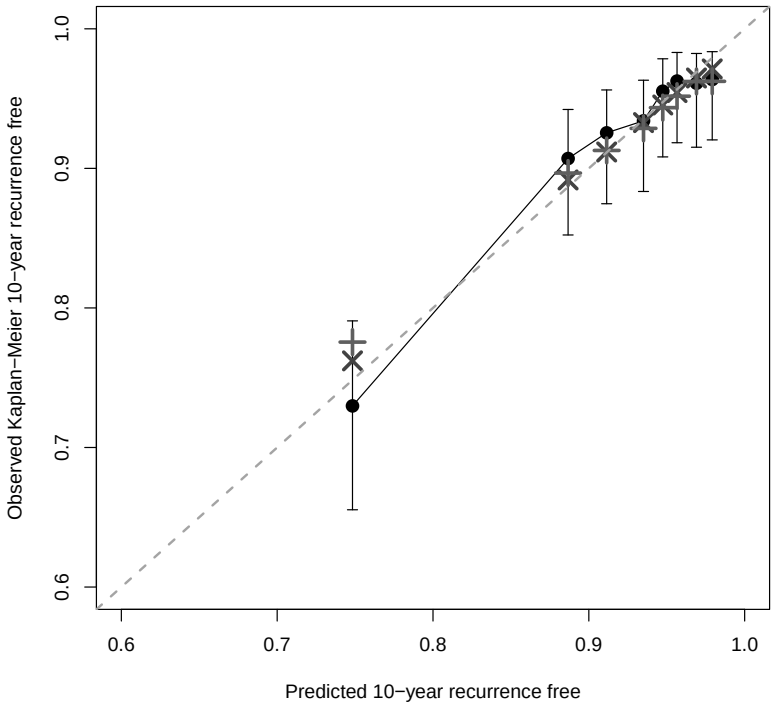


Figure 2.4: Model calibration plot. For subgroups of 200 patients, the model-predicted recurrence-free rate was plotted against the Kaplan-Meier estimated observed rates. Dots represent apparent calibration, "x" represents bootstrap corrected estimates as advised by Harrell (2001); "+" represents the modification of the bootstrap. The gray dashed line corresponds to ideal calibration.

ited or no effects and Morrow stated also "it is time to broaden our thinking beyond the simple measurement of margin width and to consider other factors that predict the likelihood of residual tumor" (Morrow, 2008; Guinot et al., 2007; Peterson et al., 1999).

As mentioned in the introduction, the IBTR! model calculates the reduction of the IBR rate after whole breast irradiation. However, their estimate of the effect of a 70% relative reduction after whole breast irradiation is based on publications from randomized trials and some retrospective data. Even in the revised IBTR! version 2.0, the impact of breast irradiation is not fully validated. We attempted to make

Table 2.4: Overview of selected variables for the final model and comparison with IBTR! version 2.0

	Final model	IBTR! version 2.0
Age	Continuous w/ spline	8 Groups
Diameter	Continuous	3 Groups
Histologic grade	2 Groups	3 Groups
Presence of DCIS	Yes/no	–
Lymphovascular invasion	(discarded)	Yes/no
Margin	(discarded)	3 Groups
Chemotherapy	Yes/no	Yes/no
Tamoxifen	Yes/no	Yes/no
Boost treatment	Yes/no	–

the same assumption of an HR of 0.3 after whole breast irradiation when using our nomogram. This would correspond to adding an extra 39 points to the patient's score in the nomogram (figure 2.3). We strongly advice to consider this only as a theoretical calculation since we could not include in our model the variable of whether whole breast irradiation was given.

We recognize some limitations to our study. The data at hand have the advantage that boost treatment was randomized, but the disadvantage that hormonal and chemotherapy treatment were not randomized. Nevertheless, there are ample data available that both hormonal and chemotherapy reduces the IBR rate, as shown in several randomized and retrospective studies. This is in line with the hazard ratios for adjuvant hormonal and chemotherapy which we found in our multivariable analysis. We realize that currently more aggressive systemic therapy is used, including trastuzumab for HER2/neu positive tumors (Mannino & Yarnold, 2009). Although nowadays shorter fractionation schedules are used for whole breast irradiation, we expect that this nomogram remains valid also for hypofractionated schedules (START Group, 2008; Whelan et al., 2010) and with a simultaneous integrated boost technique (Williamson et al., 2010; Hijal et al., 2010; Van der Laan et al., 2010). Therefore, the nomogram may not correctly evaluate today's systemic treatments. Another limitation of our study is the absence of external validation. Finally, this dataset is based on whole breast irradiation with or without a boost, and is therefore not applicable for partial breast irradiation or for complete omission of radiotherapy (Offersen et al., 2009; Polgár et al., 2010).

2.5 Summary

The nomogram (see `http://research.nki.nl/ibr`) predicting IBR after BCT includes seven factors, with young age, the presence of DCIS and boost treatment as the most dominant factors. It provides a tool to estimate the local recurrence probability after BCT and demonstrates the effect of a boost dose. Combined with our previously published model to predict fibrosis, the nomogram presented here may assist in decision making for individual patients (Collette et al., 2008).

Funding

This trial was supported by Grant No. 3U10 CA11488-18S1 through 5U10 CA011488-41 from the National Cancer Institute (Bethesda, Maryland, USA). The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute. The contribution of L. Collette to this research was supported by Fonds Cancer (FOCA), Belgium.

CHAPTER 3

Prognostic Factors for Local Control in Breast Cancer After Long-Term Follow-up in the EORTC Boost vs No Boost Trial

This chapter was published as:
Conny Vrieling*, Erik van Werkhoven*, Philippe Maingon,Philip Poortmans, Caroline Weltens, Alain Fourquet, Dominic Schinagl, Bing Oei, Carla C. Rodenhuis, Jean-Claude Horiot, Henk Struikmans, Erik van Limbergen, Youlia Kirova, Paula Elkhuizen, Rudolf Bongartz, Raymond Miralbell, David A. L. Morgan, Jean-Bernard Dubois, Vincent Remouchamps, René-Olivier Mirimanoff, Guus Hart, Sandra Collette, Laurence Collette, Harry Bartelink for the EORTC Radiation Oncology and Breast Cancer Groups
Prognostic Factors for Local Control in Breast Cancer After Long-term Follow-up in the EORTC Boost vs No Boost Trial: A Randomized Clinical Trial
JAMA Oncology 3(1):42–48, 2016
*) Equally contributing

Abstract

Importance Prognostic factors of ipsilateral breast tumor recurrence (IBTR) may change over time following breast-conserving therapy.

Objective The EORTC "boost no boost" trial showed that young age and high-grade invasive carcinoma were the most important risk factors for IBTR. This study reanalyses pathological prognostic factors related to IBTR using long-term follow-up.

Design, setting, and participants Participants included 5569 early-stage breast cancer patients, treated with breast-conserving surgery (BCS) and whole-breast irradiation (WBI), who were randomized between no boost and a 16 Gy boost in the EORTC phase III "boost no boost" trial (1989-1996). A total of 1616 patients with a microscopically complete resection (according to local pathologists), included in the central pathology review, have been analyzed in this study. Median follow-up was 18.2 years.

Interventions No further treatment or 16 Gy boost, after BCS and 50 Gy WBI.

Results The 20-year cumulative incidence of IBTR in 1616 patients (160 events observed) was 15% (95% CI, 12%-17%). Young age ($p < 0.001$) and presence of ductal carcinoma in situ (DCIS) (HR, 2.15; 95% CI, 1.36–3.38; $p = 0.001$) were associated with an increased risk of IBTR in multivariable analysis. The cumulative incidence of IBTR at 20 years was 34% (95% CI, 25%–41%), 14% (95% CI, 10%–18%), and 11% (95% CI, 8%–15%), in patients 40 years or younger, 41 to 50 years and 50 years or older, respectively ($p > 0.001$). This incidence was 18% (95% CI, 14%–22%) and 9% (95% CI, 6%–12%) for tumors with and without DCIS ($p < 0.001$). High-grade tumors relapsed more frequently early during follow-up but the relative effect of age and presence of DCIS seemed stable over time. The boost reduced the 20-year IBTR incidence from 31% (95% CI, 22%–39%) to 15% (95% CI, 8%–21%) (HR, 0.37; 95% CI, 0.22–0.62; $p < 0.001$) in high-risk patients (< 50 years with DCIS present).

Conclusions and relevance The association of high-grade invasive tumor with IBTR diminished during follow-up, while the effect of DCIS adjacent to invasive tumor seemed to remain stable. Therefore, patients with high-grade invasive tumors should be monitored closely, especially in the first 5 years, while additional DCIS is an indication for longer follow-up, emphasizing the importance of long-term trial follow-up to estimate absolute effects accurately.

Trial registration clinicaltrials.gov Identifier: NCT02295033

Key Points

Question What is the long-term impact of prognostic factors on ipsilateral breast tumor recurrence (IBTR) in patients treated with breast-conserving therapy?

Findings Young age and the presence of ductal carcinoma in situ (DCIS) adjacent to the invasive tumor were associated with an increased incidence of IBTR at long-term follow-up, whereas high-grade tumors relapsed more frequently only during the first 5 years.

Meaning Patients with high-grade invasive tumors should be monitored closely, especially in the first 5 years. The impact of DCIS remained constant over time, indicating that long-term follow-up is necessary. The boost significantly reduced IBTR in these patients.

3.1 Introduction

Since the introduction of breast-conserving therapy (BCT), several retrospective and prospective studies have analyzed clinical and pathological prognostic factors influencing local control. These studies aimed to identify clinical, radiological, and pathological criteria that would guide the individualization of surgery (mastectomy vs breast-conserving surgery [BCS]) and radiotherapy (treatment volume and dose: whole-breast irradiation [WBI] with or without a tumor bed boost vs partial breast radiotherapy or no radiotherapy at all). Well-established risk factors are first of all the conventional staging system (tumor size and nodal presence) followed by several other criteria, such as young age (Chen et al., 2015; Vrieling et al., 2003), mammographic density (Eriksson et al., 2013), margin status (Moran et al., 2014), peri-tumoral vascular invasion (Voogd et al., 2001), and molecular subtype (Lowery et al., 2012).

In the EORTC 10801 study, the long-term follow-up showed a higher local recurrence rate after BCT compared with modified radical mastectomy. Despite this result, the survival was equal in both treatment arms (Litière et al., 2012). Risk factors for local recurrence were studied combined with the Danish Breast Cancer Cooperative Group 82TM study. Young age and the presence of an extensive intraductal component (EIC) were associated with an increased risk of local recurrence after BCT. Vascular invasion was a risk factor independent of treatment. The subgroup of patients with a lobular carcinoma fared better with BCT (Voogd et al., 2001).

More recently Liu et al. (2015) showed that the intrinsic sub type of breast cancer was significantly related to the 10-year in-breast recurrence in node-negative early breast cancer patients older than 50 years, treated with tamoxifen and postoperative radiotherapy or tamoxifen alone (Bellon, 2015), varying from 5% for luminal A tumors to 21% for high-risk tumors (Her2 positive or triple-negative tumors). The subtype itself was not predictive of benefit from radiotherapy.

Earlier analyses of the EORTC boost no boost study found that young age and high-grade tumors were associated with a higher risk of local recurrence after BCT (Jones et al., 2009). With a radiotherapy boost dose of 16 Gy following WBI, the local recurrence rate could be reduced by nearly a factor of 2, resulting in the greatest absolute benefit in the youngest patients (Bartelink et al., 2015). In this trial a central pathology review was carried out in a subgroup of patients with a complete resection of the breast tumor according to the local pathologist.

In this article we reanalyze in the centrally reviewed subset the effect of pathological factors on local control with long-term follow-up, with a special focus on assessing the evolution of these effects over time. We also analyze the long-term outcome of subgroups resembling the intrinsic subtypes, and describe the effect of the radiotherapy boost in these subgroups.

3.2 Methods

Patients and Methods

The trial protocol can be found online. A total of 5569 early breast cancer patients were randomized in the EORTC boost no boost trial from 1989 to 1996. The main aim of the trial was to evaluate the influence of a boost dose in BCT in terms of local control, survival, and cosmetic outcome. The patients were treated with lumpectomy, axillary dissection, and WBI (25 times 2 Gy in 5 weeks). The 5318 patients with a microscopically complete resection according to the local pathologist were randomized between no boost and a 16 Gy boost to the tumor bed. According to the trial protocol, only patients with positive axillary lymph nodes received systemic therapy: chemotherapy for premenopausal patients and tamoxifen for postmenopausal women. Details of the trial have been published previously (Bartelink et al., 2015, 2001, 2007). Oral informed consent was obtained according to EORTC guidelines and the local and national rules of the participating institutes. Ethics committees of the participating institutes approved the protocol. Tissue blocks of 1616 patients from the first years of the accrual period underwent central pathology review, representing 30% of the overall population. Data of this subgroup with a median follow-up of 18.2 years was analyzed.

Pathology Review

The tumor characteristics and margin status were reviewed by the late breast pathologist J.L. Peterse. The extent of ductal carcinoma in situ (DCIS) was estimated by counting the number of ducts involved in the breast tissue adjacent to the primary invasive tumor. The presence of DCIS within the primary tumor was not taken into account. Up to 3 ducts involved was considered a minimal DCIS component; 4 to 9 ducts, a moderate component; and 10 or more ducts involved was considered an extensive DCIS component. Tumors consisting mainly of DCIS with focal areas of invasion were classified as invasive carcinomas with an EIC. The margin status of the invasive tumor as well as the DCIS component was defined as follows: a 'positive margin' as tumor on ink, a 'very close margin' as tumor seen at 2 mm or less from the inked resection margin, a 'close margin' as tumor seen between 3 and 4 mm and a 'free margin' as a tumor-free margin of 5 mm or more. The margin status for invasive carcinoma could be scored in 1494 patients and for DCIS in 811 patients. The histologic grade of the invasive tumor was defined according to the Elston/Ellis modification of the Bloom-Richardson system (Elston & Ellis, 1991) and the histologic grade of DCIS was classified as low, intermediate, or high (Holland et al., 1994). The subgroup of hormone-receptor negative, high-grade tumors was analyzed as surrogate for triple-negative tumors, since the Her2 status was unknown for this population. The subgroup of estrogen-receptor

positive, low-grade tumors was analyzed as surrogate for luminal A tumors.

Statistical Analysis

Time to ipsilateral breast tumor recurrence (IBTR) as first event was calculated from the date of randomization. Since it is difficult to differentiate between a local recurrence and a new primary tumor in the treated breast, all invasive recurrences found in the ipsilateral breast during follow-up were classified as IBTR. Patients alive without IBTR were censored at the date of last follow-up. Patients who first experienced another event (regional recurrence, new tumor, distant metastasis, or death) were censored at the date of this event. In addition, patients were censored if they experienced any of these other events within 4 months of their IBTR (assuming the other event was already present at the time of local recurrence), except if this concerned a regional recurrence only.

The Cox proportional hazards model was used to analyze the cause-specific hazard of IBTR, where variables included in multivariable analysis were selected based on clinical expertise and supported by univariable analysis. Interactions with time were assessed by the Pearson product-moment correlation between the scaled Schoenfeld residuals of the Cox model and log(time). A global test for interactions was significant ($p = 0.002$). For a visual inspection of possible interactions with time, the residuals were plotted against time along with a smooth curve (Therneau & Grambsch, 2000). A restricted cubic spline with 3 knots was used for age. Kaplan-Meier estimates of cumulative incidence were reported at 20 years, or at 15 years for subgroups with fewer than 20 subjects at risk at 20 years. Cox models with interactions were used to compare the effect of boost treatment between subgroups. Subjects with missing data necessary for analysis were removed from that particular analysis. Results with a p value < 0.01 were considered statistically significant.

3.3 Results

The median age of the patients was 54 years (table 3.3 on page 38). After lumpectomy and 50 Gy WBI, no boost was given in 801 patients, while 815 patients received a 16 Gy boost (figure 3.1). The median tumor size was 15 mm, most of the tumors were hormone receptor positive and 78% of patients had negative axillary lymph nodes. Patients with axillary lymph node involvement received adjuvant systemic treatment: 16% of premenopausal patients received chemotherapy and 23% of postmenopausal patients received tamoxifen (20 mg per day for 2 years). The majority of tumors were invasive ductal carcinomas, in 58% of patients associated with a DCIS component (table 3.3).

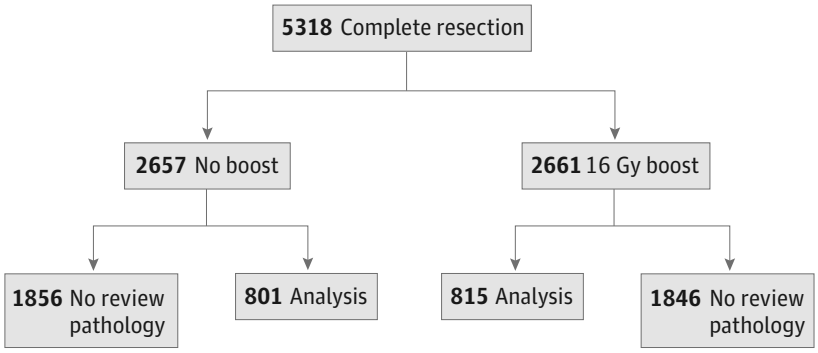


Figure 3.1: Trial population

Ipsilateral Breast Tumor Recurrence

A total of 160 IBTR as first event were found, 99 in the no-boost arm and 61 in the boost arm. The 20-year cumulative risk of IBTR was 17% (95% CI 13%–20%) and 12% (95% CI 9%–16%), respectively ($p < 0.001$) (table 3.1). The patient characteristics of the subgroup with central pathology review did not differ significantly from the population without review (Jones et al., 2009), and neither did local control. The cumulative incidence curves of IBTR never reached a plateau and the favorable absolute effect of the boost increased over time as the curves continued to diverge (figure 3.4 on page 34).

In the univariable analysis, the boost treatment and use of tamoxifen were significantly associated with improved local control, whereas young age and the presence of DCIS were prognostic of increased risk of IBTR (table 3.4 on page 39). Patients with high invasive grade were at greater hazard of IBTR in the first 5 years of follow-up, but the effect declined in the course of time (interaction with time $p < 0.001$, figure 3.2a), with more than 250 high-grade patients at risk after 5 years (figure 3.5 on page 35). For the presence of additional DCIS no such change in hazard over time was observed (interaction with time $p = 0.41$, figure 3.2b). In patients receiving systemic therapy, the boost still had a significant influence on IBTR (figure 3.6 on page 36). Neither incomplete resection of invasive tumor nor tumor-free margin distance in millimeters for complete resection was significantly related to local control. Also for the additional DCIS component, tumor-free margin in millimeters for complete resection or even incomplete resection did not appear to influence local control.

Table 3.1: Cumulative probability of ipsilateral breast tumor recurrence as first event at 20 years of follow-up (univariable effects)

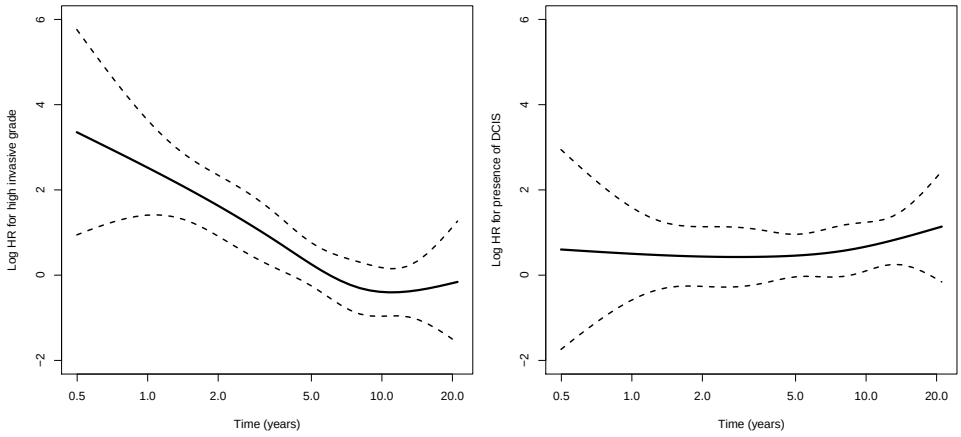
			IBTR Probability	
	Subjects	Events	% (95% CI)	p-value
Treatment				< 0.001
No boost	801	99	17 (13–20)	
16 Gy boost	815	61	12 (9–16)	
Age (years)				< 0.001
27–40	183	49	34 (25–41)	
41–50	442	44	14 (10–18)	
>50	991	67	11 (8–15)	
Presence of DCIS				< 0.001
No	664	44	9 (6–12)	
Yes	914	110	18 (14–22)	
Histological grade of invasive tumor				0.08
Low	784	70	12 (10–15)	
Intermediate	398	35	14 (9–18)	
High	363	42	16 (10–22)	

DCIS: ductal carcinoma in situ

Risk Factors for Local Recurrence

After adjustment for treatment and known prognostic factors, young age ($p < 0.001$) and presence of DCIS (HR 2.15; 95% CI 1.36–3.38; $p = 0.001$) were statistically significant predictors of IBTR (table 3.2). The histological grade of the invasive tumor did not significantly influence long-term local control. The association between age at randomization and IBTR was nonlinear (figure 3.7 on page 37), but similar in both treatment arms. The risk of IBTR decreased from age 30 to about 50 from 34% (95% CI 25%–41%) to 11% (95% CI 8%–15%) (Table 1). As of the age of 50, the risk more or less stabilized. In tumors with and without additional DCIS, the cumulative incidence of IBTR at 20 years was 18% (95% CI 14–22) and 9% (95% CI 6%–12%), respectively ($p < 0.001$, table 3.1).

A total of 124 patients had estrogen receptor (ER) and progesterone receptor-negative, high-grade tumors. In this group were 16 events. The 15-year cumulative incidence of IBTR in this population was 16% (95% CI 8%–23%). The IBTR incidence related to age showed the following trend: 15-year cumulative incidence of IBTR was 34% (95% CI 9%–53%) in patients younger than 40 years of age, 19% (95% CI 2%–32%) in patients aged 41 to 50, compared with 6% (95% CI 0%–12%) for patients older than 50 years ($p = 0.04$). The presence of additional DCIS did



(**a**) High invasive grade (**b**) Ductal carcinoma in situ (DCIS)

Figure 3.2: Log HR over time for (a) high invasive grade ($p < 0.001$) and (b) presence of DCIS ($p = 0.41$) with confidence intervals at 2 standard errors. The time axis has been graduated according to the log scale.

not influence local control in this population.

464 patients had ER-positive, low-grade tumors. In this group were 43 events. The 15-year cumulative incidence of IBTR in this subgroup was 11% (95% CI 8%–14%). Age had a significant influence also in this population: patients younger than 40 years had a 15-year IBTR incidence of 34% (95% CI 14%–49%) compared with 10% (95% CI, 5%–15%) for patients 40 years or older ($p < 0.001$). The presence of DCIS in this population showed a trend in 15-year IBTR incidence: 14% (95% CI 9%–19%) for patients with additional DCIS vs 7% (95% CI, 3%–11%) for patients without ($p = 0.02$).

Effect of the Boost Treatment in High-Risk Patients

The influence of the radiotherapy boost on the different subgroups is shown in a forest plot (figure 3.3).

For patients younger than 50 years, the 16 Gy boost dose reduced the 20-year cumulative incidence of IBTR from 24% (95% CI 18%–30%) to 15% (95% CI 10%–20%) (HR 0.51; 95% CI 0.33–0.77; $p = 0.002$). In patients with additional DCIS, the boost dose reduced the 20-year cumulative incidence of IBTR from 22% (95% CI 17%–27%) to 14% (95% CI 9%–19%) (HR 0.47; 95% CI 0.31–0.69; $p < 0.001$). In the population with both risks combined, the boost dose reduced the 20-year cumulative incidence of IBTR from 31% (95% CI, 22%–39%) to 15% (95% CI 8%–21%) (HR

Table 3.2: Multivariable analysis for ipsilateral breast tumor recurrence as first event

Variable	HR (95% CI)	p-value
Treatment		
No Boost vs 16 Gy Boost	0.62 (0.41–0.92)	0.02
Age		
Per year	(See figure 3.7)	< 0.001
Positive nodes		
No vs Yes	0.82 (0.43–1.56)	0.55
Systemic therapy*		
No vs Yes	0.76 (0.44–1.29)	0.31
Diameter		
Per mm	1.03 (1.00–1.06)	0.05
Grade invasive tumor		
Intermediate/low vs High	0.87 (0.52–1.46)	0.60
DCIS		
No vs Yes	2.15 (1.36–3.38)	0.001
Estrogen		
Negative vs Positive	1.11 (0.67–1.85)	0.67
Progesterone		
Negative vs Positive	0.79 (0.48–1.29)	0.34

DCIS: ductal carcinoma in situ

*) Systemic therapy indicates tamoxifen or chemotherapy

0.37; 95% CI, 0.22–0.62; $p < 0.001$). The influence of the boost in the older patients with DCIS (545 patients with 45 events) was not significant: a 20-year cumulative incidence of IBTR of 15% (95% CI 9%–21%) without vs 14% (95% CI 5%–23%) with the boost ($p = 0.11$).

For the subgroup of patients with hormone receptor-negative, high-grade tumors, the 16 Gy boost dose reduced the 15-year cumulative incidence of IBTR from 31% (95% CI 14%–44%) to 5% (95% CI, 0%–9%) (HR 0.23; 95% CI 0.07–0.70; $p = 0.01$).

For patients with ER-positive, low-grade tumors, the 16 Gy boost dose did not change the IBTR rate. Neither was this the case for patients with ER positive, low-grade tumors and additional DCIS.

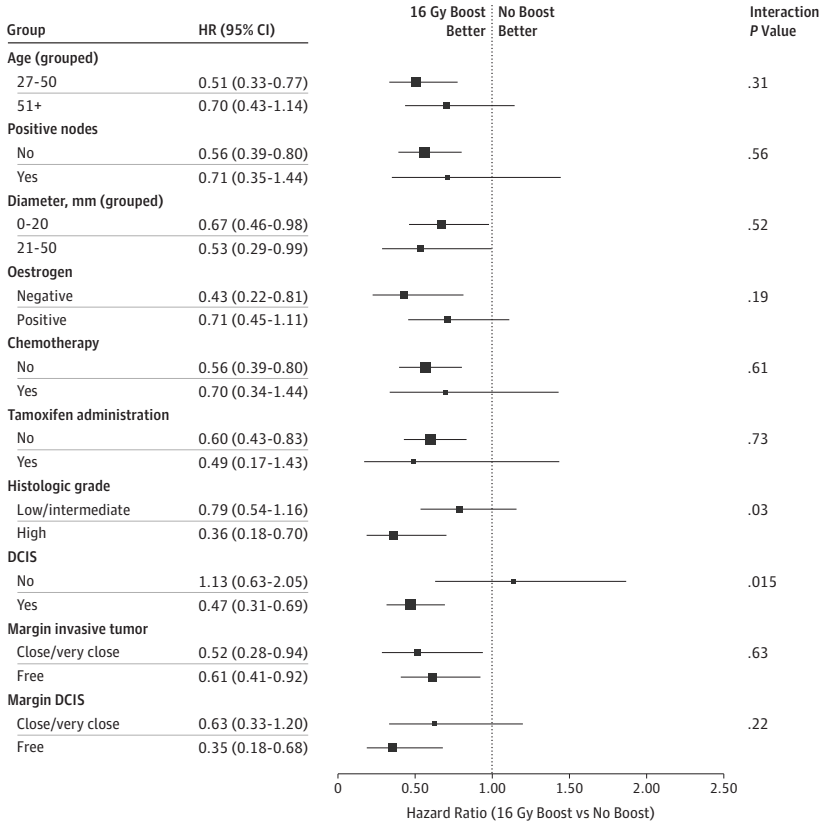


Figure 3.3: Effect of the Radiotherapy boost dose on IBTR for the different subgroups

3.4 Discussion

This long-term analysis of randomized BCT patients with a central pathology review showed that 2 factors had a significant negative impact on local control: young age and the presence of an additional DCIS component adjacent to the primary tumor.

The negative impact of young age on local control could be reduced by the boost in radiation therapy dose. Bartelink et al. (2015) showed that the relative improvement in local control by the boost was similar for the different age groups; however, the absolute risk reduction in local recurrence was the largest in the younger patients. Younger patients were not at higher risk for adverse effects of the boost dose in terms of cosmetic outcome or fibrosis development, which remained inde-

pendent from age (Collette et al., 2008).

In our previous analysis (Jones et al., 2009), we showed that the grade of invasive cancer, together with boost and age, remained significant in the final multivariable analysis. With longer follow-up, the relative effect of invasive tumor grade decreased rapidly within the first 5 years, losing its significance with longer follow-up, whereas the relative effect of presence of DCIS did not diminish over time (figure 3.2), doubling the IBTR incidence at 20 years. This factor has a constant relative effect on local control over time, meaning that the absolute difference between the cumulative incidence curves continues to widen, emphasizing the importance of long-term follow-up (Pritchard & Sousa, 2011).

All excisions were complete according to local pathologists, but resection margin width was analyzed from central review data. Neither the margin for the invasive tumor nor for the associated DCIS component was associated with IBTR. These results confirm the Society of Surgical Oncology–American Society for Radiation Oncology consensus guidelines on margins in BCS for invasive cancer (Moran et al., 2014). These guidelines concluded that positive margins (ink on tumor) were associated with a 2-fold increase in the risk of local recurrence, but if the margins were negative (no ink on tumor) an increase in margin width did not significantly decrease the risk of local recurrence.

Based on the risk factors for local control, we defined high- and low-risk populations. The high-risk group consisted of patients 50 years or younger with DCIS in addition to the invasive tumor, in which the boost dose reduced the incidence of IBTR with a HR of 0.37 (95% CI 0.22–0.62) translating in an absolute decrease of 16% at 20 years.

A low-risk group was defined as patients having ER-positive, low-grade tumors (as an approach for the selection of luminal A tumors). The radiotherapy boost did not appear to modify the risk of local relapse in this subgroup. We know that overall the percentage of local recurrences in early breast cancer patients is decreasing (Bartelink et al., 2012; Poortmans et al., 2012). In this favorable population, the question is whether they need any radiotherapy at all (Lowery et al., 2012; Voduc et al., 2010). Three different studies randomized postmenopausal women with low-risk hormone receptor-positive early breast cancer treated with BCS and endocrine therapy between WBI and no further treatment (Hughes et al., 2004; Fyles et al., 2004; Kunkler et al., 2015). The 5-year results show an IBTR incidence of 0.6% to 1.3% in the WBI group compared with 4.0% to 7.7% in case of no RT. Only Hughes et al. (2013) published the long-term follow-up results: a 10-year loco-regional recurrence rate of 2% in the WBI group compared with 10% in the no RT patients. This result underlines the need for long-term follow-up given the pattern of late recurrences in these favorable tumors (Ribelles et al., 2013). Liu et al. (2015) concluded that patients older than 60 years with T1 luminal A tumors, treated with

lumpectomy and tamoxifen alone, had a 10-year IBTR of only 3.1%. Currently, several single-arm trials of BCS and endocrine therapy without radiotherapy in postmenopausal patients with small luminal A tumors are initiated (Clinicaltrials.gov NCT01791829, NCT02400190, NCT02653755). The development of gene-expression signatures related to local control in breast cancer is another important tool in the selection of patients benefiting from postoperative radiotherapy (Tramm et al., 2014), 27 but a reliable and validated profile predicting the need for postoperative radiotherapy is currently not yet available.

Limitations

There are limitations to this study. The pathology review population was limited to less than one-third of the whole trial population. Therefore, the subgroup analysis does not have much power, and although the forest plot indicates homogeneity of the effect of radiotherapy boost treatment, the nonsignificance of interactions should be interpreted with caution. Furthermore, the IBTR rates have fallen greatly in the past years, so the absolute risk reduction caused by the boost is currently probably smaller. Owing to the absence of a treatment arm without radiotherapy, we could not study the possibilities of omitting radiotherapy for favorable subgroups. As the measurement of HER2/neu was not standard during the course of the trial, we were unable to fully assess the impact of subtyping on local control.

3.5 Conclusions

The long-term follow-up analysis of pathological prognostic factors associated with local control in the EORTC boost no boost trial showed that young age and the presence of associated DCIS increase the risk of IBTR. In patients with both factors the radiotherapy boost dose reduced the IBTR risk with an HR of 0.37, leading to an absolute risk reduction of 16% at 20 years. The proportional hazards assumption of a constant hazard was valid for almost all variables, except for the effect of high histologic grade, which diminished over time. The fact that the relative impact of additional DCIS on local control seemed to remain constant over time, whereas the impact of high grade decreased over time, underlines the importance of long-term trial follow-up to correctly estimate absolute effects. Patients with high-grade invasive tumors need to be monitored closely especially in the first 5 years, whereas patients with invasive tumors with associated DCIS need long-term follow-up, at least 20 years.

3.6 Supplementary Material

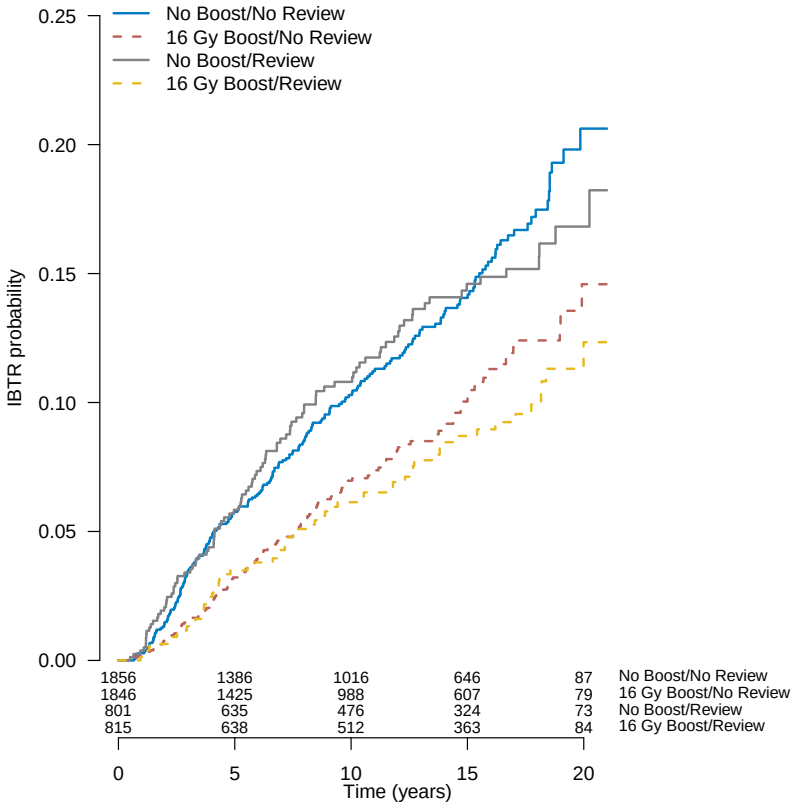


Figure 3.4: Cumulative incidence of IBTR according to the boost treatment in the review population compared to the population without review

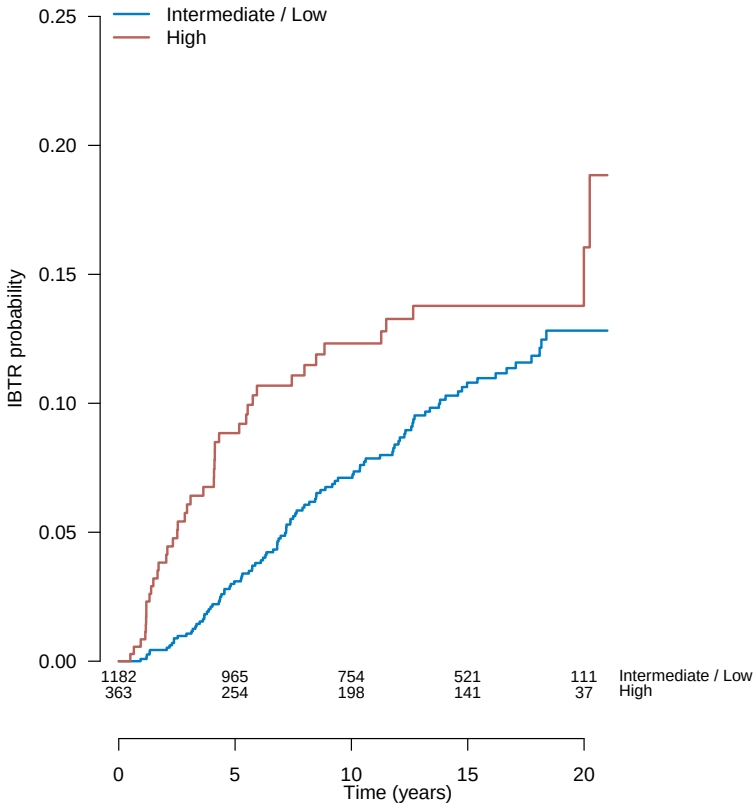


Figure 3.5: Cumulative incidence of IBTR by histologic grade of invasive tumor

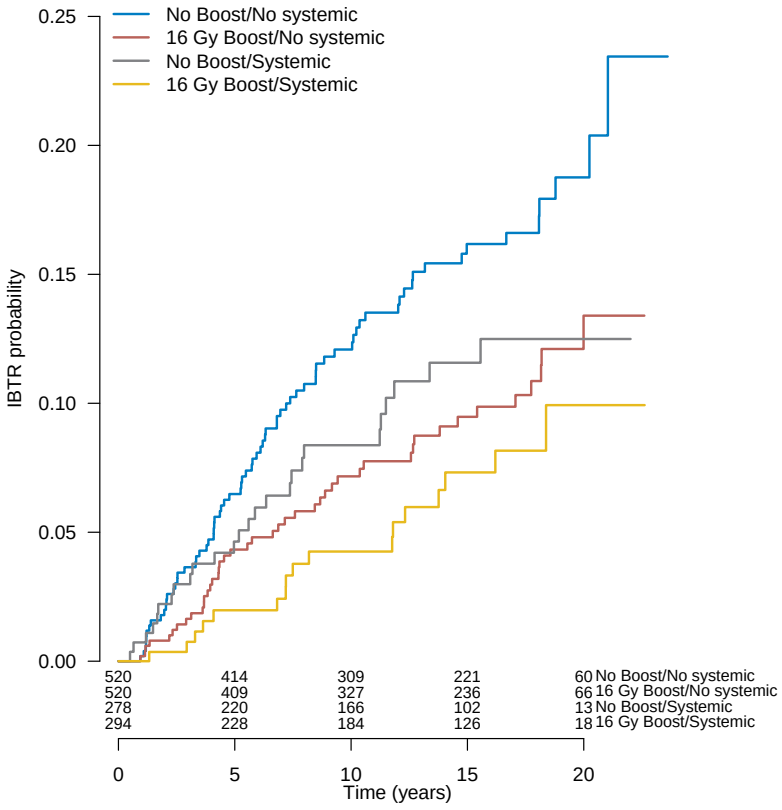


Figure 3.6: Cumulative incidence of IBTR by boost and systemic treatment

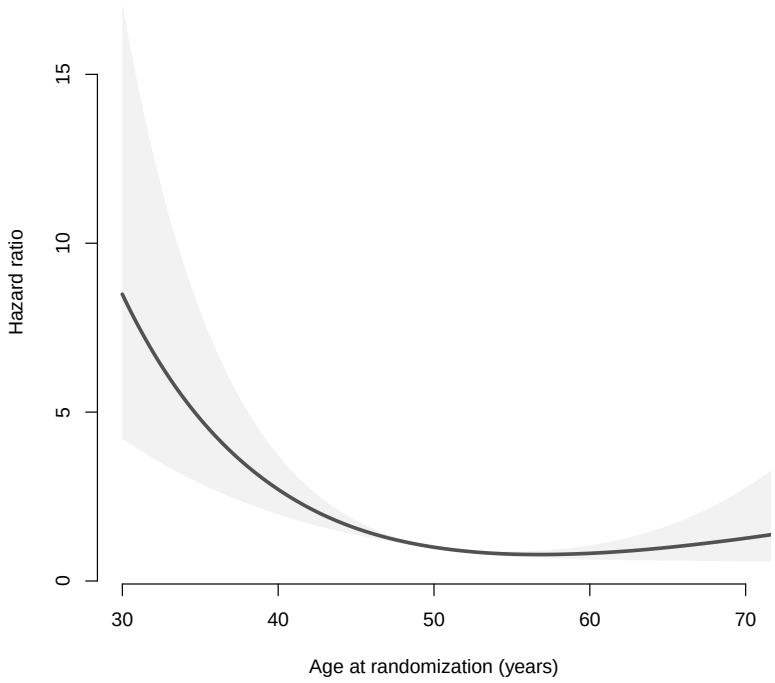


Figure 3.7: Spline curve for the effect of age in the multivariable model for IBTR

Table 3.3: Baseline patient and tumor characteristics

	Randomised Treatment		
	No boost	Boost	Total
	801	815	1,616
Age at randomization (years)			
Median (range)	55 (28–75)	54 (27–76)	54 (27–76)
Diameter of dominant lesion (mm)			
Median (range)	15 (2–50)	15 (0–50)	15 (0–50)
Estrogen			
Negative	165 (28%)	172 (29%)	337 (28%)
Positive	420 (72%)	429 (71%)	849 (72%)
NA	216	214	430
Progesterone			
Negative	187 (36%)	193 (35%)	380 (36%)
Positive	329 (64%)	355 (65%)	684 (64%)
NA	285	267	552
Positive nodes			
No	624 (79%)	619 (76%)	1,243 (78%)
Yes	169 (21%)	191 (24%)	360 (22%)
NA	8	5	13
Tamoxifen			
No	616 (77%)	625 (77%)	1,241 (77%)
Yes	182 (23%)	189 (23%)	371 (23%)
NA	3	1	4
Chemotherapy			
No	672 (84%)	687 (84%)	1,359 (84%)
Yes	126 (16%)	127 (16%)	253 (16%)
NA	3	1	4
Systemic therapy*			
No	520 (65%)	520 (64%)	1,040 (65%)
Yes	278 (35%)	294 (36%)	572 (35%)
NA	3	1	4
Histology			
Ductal	545 (70%)	587 (73%)	1,132 (71%)
Lobular	45 (6%)	48 (6%)	93 (6%)
Mixed pattern	87(11%)	78 (10%)	165 (10%)
Other	106 (14%)	94 (12%)	200 (13%)
NA	18	8	26
Histological grade of invasive tumor			
Low	374 (49%)	410 (53%)	784 (51%)
Intermediate	217 (28%)	181 (23%)	398 (26%)
High	174 (23%)	189 (24%)	363 (23%)
NA	36	35	71
Margin of invasive tumor			
Free	453 (62%)	497 (66%)	950 (64%)
Close	96 (13%)	91 (12%)	187 (13%)
Very close	160 (22%)	146 (19%)	306 (20%)
Positive	27 (4%)	24 (3%)	51 (3%)
NA	65	57	122

(Continued on next page)

Table 3.3 – continued from previous page

	No boost 801	Boost 815	Total 1,616
Presence and type of CIS			
None	308 (40%)	291 (36%)	599 (38%)
Ductal	416 (53%)	449 (56%)	865 (55%)
Lobular	29 (4%)	36 (4%)	65 (4%)
Both ductal and lobular	25 (3%)	24(3%)	49 (3%)
NA	23	15	38
Histological grade of DCIS			
Low	112 (25%)	110 (23%)	222 (24%)
Intermediate	195 (44%)	210 (45%)	405 (44%)
High	133 (30%)	151 (32%)	284 (31%)
No DCIS	337	327	664
NA	24	17	41
Presence of EIC			
No	328 (78%)	354 (80%)	682 (79%)
Yes	90 (22%)	89 (20%)	179 (21%)
No DCIS	337	327	664
NA	46	45	91
Margin of DCIS			
Free	164 (42%)	193 (46%)	357 (44%)
Close	51 (13%)	69 (16%)	120 (15%)
Very close	109 (28%)	109 (26%)	218 (27%)
Positive	63 (16%)	53 (12%)	116 (14%)
No DCIS	337	327	664
NA	77	64	141

*) tamoxifen and/or chemotherapy
CIS: carcinoma in situ, DCIS: ductal carcinoma in situ
EIC: extensive intra-ductal component

Table 3.4: Univariable analysis of local relapse as first event

	Patients	Events	HR	95% CI	p-value
Age (per year)	1,616	160	0.94	0.93–0.96	< 0.001
Diameter dominant lesion (per mm)	1,591	154	1.02	1.00–1.04	0.07
Estrogen					
Negative	337	42	1		
Positive	849	79	0.7	0.48–1.02	0.07
Progesterone					
Negative	380	42	1		
Positive	684	67	0.84	0.57–1.23	0.37
Positive nodes					
No	1,243	129	1		
Yes	360	31	0.89	0.60–1.32	0.56

(Continued on next page)

Table 3.4 – continued from previous page

	Patients	Events	HR	95% CI	p-value
Randomized treatment					
No boost	801	99	1		
16 Gy boost	815	61	0.59	0.43–0.81	0.001
Tamoxifen					
No	1,241	145	1		
Yes	371	15	0.37	0.22–0.64	< 0.001
Chemotherapy*					
No	1,359	130	1		
Yes	253	30	1.21	0.81–1.80	0.35
Systemic therapy					
No	1,040	117	1		
Yes	572	43	0.7	0.49–0.99	0.04
Histology ($p = 0.71$, df=3)					
Ductal	1,132	109	1		
Lobular	93	8	0.85	0.41–1.74	0.66
Mixed pattern	165	13	0.76	0.43–1.36	0.36
Other	200	23	1.1	0.70–1.72	0.69
Differentiation grade of invasive tumor ($p = 0.09$, df=2)					
Low	784	70	1		
Intermediate	398	35	1.11	0.74–1.67	0.61
High	363	42	1.54	1.05–2.26	0.03
Margin of invasive tumor ($p = 0.14$, df=3)					
Free	950	95	1		
Close	187	20	1.1	0.68–1.78	0.70
Very close	306	29	0.97	0.64–1.47	0.89
Positive	51	1	0.19	0.03–1.34	0.10
Presence and type of CIS ($p = 0.003$, df=3)					
None	599	39	1		
Ductal	865	100	1.78	1.23–2.57	0.002
Lobular	65	5	1.15	0.45–2.92	0.77
Ductal and lobular	49	10	2.93	1.46–5.86	0.003
Differentiation grade of DCIS ($p = 0.21$, df=2)					
Low	222	27	1		
Intermediate	405	41	0.91	0.56–1.48	0.74
High	284	42	1.34	0.82–2.17	0.24
EIC					
No	682	82	1		
Yes	179	24	1.06	0.67–1.67	0.80
Margin DCIS ($p = 0.78$, df=3)					
Free	357	43	1		
Close	120	12	0.80	0.42–1.53	0.51
Very close	218	26	1.11	0.68–1.81	0.67
Positive	116	16	1.12	0.63–1.99	0.69

CIS: carcinoma in situ, DCIS: ductal carcinoma in situ
EIC: extensive intra-ductal component
*) only a small subgroup of patients received adjuvant chemotherapy based on poor prognostic factors

CHAPTER 4

Factors Associated With Patient-Reported Cosmetic Outcome in the Young Boost Breast Trial

This chapter was published as:
Patricia J.A.M. Brouwers*, Erik van Werkhoven*, Harry Bartelink, Alain Fourquet, Claire Lemanski, Judith van Loon, John H. Maduro, Nicola S. Russell, Luc J. E. E. Scheijmans, Dominic A. X. Schinagl, Antonia H. Westenberg, Philip Poortmans, Liesbeth J. Boersma on behalf of the Young Boost Trial research group
Factors associated with patient-reported cosmetic outcome in the Young Boost Breast Trial
Radiotherapy and Oncology 120(1):107–113, 2016.
*) Equally contributing

Abstract

Purpose To investigate which factors are related to patient reported cosmetic outcome (PRCO) after breast conserving therapy.

Methods From 2004 to 2011, 2421 cT1-2N0-2a breast cancer patients were randomised in the Young Boost Trial between a 16 and a 26 Gy boost to the tumour bed. Cosmesis was scored subjectively by the patient and physician, and objectively using BCCT.core, at baseline, one and four years after treatment. Presence of fibrosis, quality of life (QoL) and rib pain at four years were also scored. Data were complete for 864 patients. The relation between the separate components was investigated using a proportional odds model.

Results Of the 7 BCCT.core parameters, the distance from nipple to inframammary fold and the length of the breast contour were significantly related to the overall PRCO at four years. Patients with more fibrosis and poorer QoL scored their cosmesis worse, while rib pain was not related. The agreement between the different scores was low (κ 0.26–0.42).

Conclusions The distance from nipple to inframammary fold, the length of the breast contour and the severity of fibrosis were the main factors related to patient-reported cosmetic outcome. Patients with better QoL scored their cosmesis better.

4.1 Introduction

The EORTC boost-no boost trial showed that adding a 16 Gy boost to the primary tumour bed after 50 Gy whole breast irradiation, reduces the local recurrence rate (LRR) with 35% (Bartelink et al., 2015). Nevertheless, even after a boost, the LRR in young patients (≤ 50 years of age) remained higher than 1% per year. Therefore, in 2004, the Young Boost trial (YBT) was launched (NCT00212121), with the primary aim to investigate whether a higher boost dose of 26 Gy would further reduce the LRR in young patients. Since the boost-no boost trial showed that the boost led to a worse cosmetic outcome (Vrieling et al., 1999), cosmetic outcome was an important secondary endpoint in the YBT.

Scoring cosmesis is difficult and often considered as controversial, because of its subjective nature. For example: Mukesh et al. (2013b, 2014) found that physicians judged cosmetic outcome to be superior after Intensity Modulated Radiotherapy (IMRT) compared to 2D radiotherapy, whereas the patient reported cosmetic outcome (PRCO) showed no benefit of IMRT. A recent analysis of the START trials showed that despite a low agreement between different scoring methods of cosmetic outcome, each scoring method could sufficiently discriminate different fractionation schedules (Haviland et al., 2016). In most studies different scoring methods are reported, including patient questionnaires, scoring by professionals (or a panel) and/or a photographic assessment using objective and reproducible software programs, such as BCCT.core (Cardoso et al., 2007) or BAT (Fitzal et al., 2007).

Although the objective methods seem to be the most attractive due to their good reproducibility, they are mainly based on measures to quantify asymmetry, assuming that symmetry is the most important determinant for PRCO. However, if that were true, a much better correlation between PRCO and objective measures would be expected than described in literature. We hypothesised that specific aspects of symmetry (e.g. nipple position) are more important for patients than other aspects (e.g. breast size), and that other factors such as pain or palpable firmness of the breast also influence PRCO. The aim of the current paper was therefore to prospectively investigate which objective cosmetic factors are associated with PRCO in the YBT. We also analysed the relation between fibrosis, pain and quality of life (QoL) with PRCO.

4.2 Patients and methods

Patient population and treatment

Patients younger than 51 years with non-metastatic, histologically proven invasive breast cancer, pT1-2pN0-2a (Sobin et al., 2002), with an Eastern Cooperative Oncology Group (ECOG) performance scale ≤ 2 (Young et al., 2015), were eligible for

the trial. Tumours were completely removed by wide local excision, although focally involved margins were allowed, defined as: 'tumour (ductal carcinoma in situ or invasive carcinoma) on ink in an area of less than 4 mm'. Sentinel lymph node biopsy and/or axillary lymph node dissection had to be performed. No neoadjuvant systemic treatment was allowed. No previous history of malignant disease, except adequately treated carcinoma in situ of the cervix or basal cell carcinoma of the skin was allowed.

Patients were randomised to a standard 16 Gy or a high 26 Gy boost to the tumour bed after 50 Gy whole breast irradiation. Other fractionation schemes, including simultaneous integrated boost techniques were allowed as well, as long as the biologically equivalent dose, calculated with an α/β of 10 for tumour, was similar. Stratification factors were age ($\leq$ vs. $>$ 40 yr), pathological tumour size ($\leq$ vs. $>$ 3 cm), oestrogen receptor status, nodal status, interstitial/external boost and institute. Patients were stratified at the time of randomisation using a minimisation technique. The study was centrally approved by the medical ethical committee of the Netherlands Cancer Institute and by the local medical ethics committees. All patients gave their written informed consent to participate. The study was registered at Clinicaltrials.gov as NCT00212121.

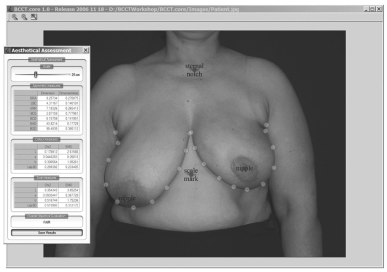
Cosmetic outcome

BCCT.core software (Cardoso et al., 2007; Cardoso & Cardoso, 2007) Digital photographs in anterior-posterior view were analysed using the BCCT.core software program, resulting in an objective score for the overall cosmetic outcome: excellent, good, fair or poor. This score is based on symmetry, skin colour and scar visibility (figure 4.1a on page 45). The seven features of symmetry in the BCCT.core program are:

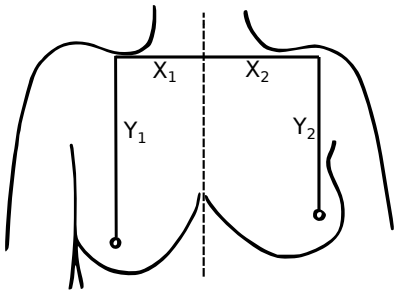
- breast retraction assessment (BRA)
- level of lower breast contour (LBC)
- upward nipple retraction (UNR)
- breast compliance evaluation (BCE; distance from nipple to inframammary fold)
- breast contour difference (BCD)
- breast area difference (BAD)
- breast overlap difference (BOD)

For all symmetry features a relative value was calculated by the program resulting in a pBRA, pLBC etcetera (examples in figures 4.1b–4.1d).

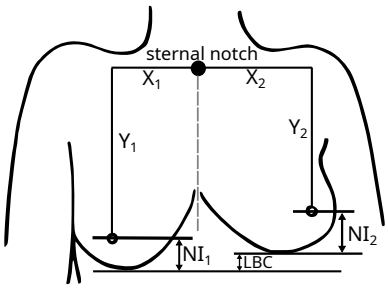
Physician's score Physicians scored using the Harris scale on overall cosmetic outcome: excellent, good, fair or poor (Harris et al., 1979).



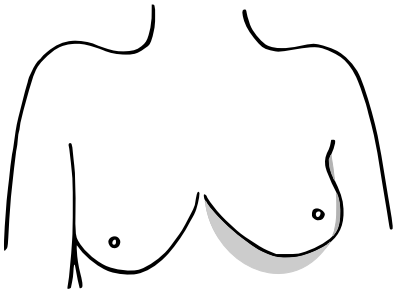
(**a**) BCCT.core software



(**b**) Breast Retraction Assessment $BRA = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$
$pBRA = \frac{2 \times BRA}{\sqrt{X_1^2+Y_1^2}+\sqrt{X_2^2+Y_2^2}}$



(**c**) Lower Breast Contour (LBC) $LBC = |(Y_1 + NI_1) - (Y_2 + NI_2)|$
$pLBC = \frac{2 \times LBC}{Y_1+NI_1+Y_2+NI_2}$



(**d**) Breast Overlap Difference (BOD) (non-overlapping area of the breasts)
$pBOD = \frac{2 \times BOD}{\text{right area}+\text{left area}}$

Figure 4.1: Examples of BCCT.core parameters, including formulas for the relative value

Patient's questionnaire The PRCO was determined by asking patients to complete the questionnaire developed by Sneeuw et al. (1992). In this validated questionnaire (page 54) overall cosmetic outcome was rated on a five-point scale: very satisfied, satisfied, not dissatisfied, dissatisfied and very dissatisfied. The patients were also asked to rate the difference between the treated breast and the untreated breast in terms of scar visibility, difference in size, shape, colour, nipple position, and firmness on a four-point scale: no difference, small difference, quite a lot of difference, or a large difference.

Other variables At the same time points fibrosis (whole breast) was scored by the physician on a four-point scale. The presence of rib pain was scored separately (yes/no). At four years, quality of life (QoL) was scored using the EORTC QLQ C-

30 questionnaire (Aaronson et al., 1993). The global QoL was measured on a scale from 1 to 7. Emotional functioning was measured on a multi-item scale ranging from 0 to 100. The parameter value was calculated for a difference of 10 points. Depression was measured at a scale from 1 to 4. A higher score on the functional scale and global QoL implies better score, while a higher score on the depression scale implies more symptoms.

Analysis First, we analysed the correlation of overall cosmetic outcome between the three scoring methods, and between fibrosis scored by the physician and firmness of the breast scored by the patient.

Secondly, we analysed the seven features of BCCT.core in a proportional odds model, to investigate which parameters were related to the PRCO at four years. Also, we analysed whether fibrosis, presence of rib pain or QoL was related to the PRCO.

To evaluate the correlation between the different factors and overall cosmetic outcome, we defined two categories: satisfactory overall cosmetic outcome and unsatisfactory overall cosmetic outcome. Excellent and good as well as very satisfied and satisfied were grouped as 'satisfactory'; fair and poor, not dissatisfied, dissatisfied, and very dissatisfied were grouped as 'unsatisfactory'.

Statistics Agreement between the three different scoring systems was calculated by Cohen's kappa statistics. The kappa coefficient (κ) is a common measure for agreement (Cohen, 1968). The overall cosmetic outcome was evaluated on a five-point scale by the patient's questionnaire but on a four-point scale by the BCCT.core software and physician. Therefore, the agreement of the overall cosmetic outcome was assessed using the grouped dichotomised outcome variable as described above. For the agreement on individual (separate) cosmetic outcome parameters, all three used a four point scale and therefore a weighted kappa ($w\kappa$) was used, where the weights were chosen quadratic. A value of 0–0.2 for κ indicates a slight agreement, 0.2–0.4 indicates a fair agreement, 0.4–0.6 indicates a moderate agreement, 0.6–0.8 indicates a substantial agreement and a value of 0.8–1.0 indicates an almost perfect agreement.

Associations between PRCO and the seven BCCT.core parameters were assessed with proportional odds models, taking into account the ordinal nature of the outcome. For each type a higher score means a worse outcome. An important assumption of this cumulative link model is that the association between each pair of outcome groups is the same, so that for example the comparison between a score of 1 versus a score of 2, 3 or 4, and the comparison of 1 or 2 versus 3 or 4 can be modelled by the same parameter. This is called the proportional odds assumption. To assess whether pain, fibrosis or QoL parameters were associated with worse PRCO,

these were analysed in models where the BCCT.core outcome was entered as a covariate. The adjustment for BCCT.core outcome gives the parameters for fibrosis, rib pain, or QoL the interpretation of what the difference would be between two patients with the same BCCT.core outcome who differ only in their fibrosis, rib pain or QoL. The QoL parameters were entered as a continuous variable in the model.

4.3 Results

Between 2004 and 2011, 2421 breast cancer patients were included in 18 institutes from The Netherlands, 13 institutes from France and 1 institute from Germany. 1211 patients were randomised to receive a standard 16 Gy boost and 1210 patients to a high 26 Gy boost. Median age was 45 years (range 19–51), 19% was younger than 40 years old. 72% of patients had a T1 tumour and 28% of patients had a T2 tumour. 61% of patients received adjuvant chemotherapy and 39% did not receive adjuvant chemotherapy. Median follow-up at the time of this analysis was 51 months. At four years we had evaluable digital photographs of 805 patients, of whom 684 also had an evaluable photograph at baseline. 1204 patients filled in the questionnaire at baseline, of whom 697 filled one in at four years too. The cosmetic result was scored by the physician for 1914 patients at baseline, and for 864 at both baseline and 4 years (figure 4.2).

Overall cosmetic outcome for the different scoring systems, and correlation between scoring systems

At four years, the BCCT.core program yielded a 'satisfactory' (i.e. excellent or good) overall cosmetic outcome in 61% of patients. The physicians and patients scored the overall cosmetic outcome as 'satisfactory' in 56% and 57% of patients, respectively. The agreement between the physician and the patient scores was moderate ($\kappa = 0.42$), between the patient and BCCT.core fair, and between the physician and BCCT.core scores the agreement was fair, with κ values of 0.26 and 0.39, respectively. The agreement between firmness scored by the patient and the grade of fibrosis scored by the physician was fair ($w\kappa = 0.36$, 95% CI 0.29–0.42) (table 4.1).

Objective factors associated with patient reported cosmetic outcome

Of the seven BCCT.core parameters, pBCE (distance from nipple to inframammary fold) and pBCD (length of breast contour) were significantly associated with a worse PRCO at four years (table 4.2). Patients with fibrosis had worse PRCO than patients without fibrosis, even when the objective score based on BCCT.core was similar (i.e., after adjustment for it). The same was true for difference in firmness

Figure 4.2: Flow diagram of available and evaluable digital photographs, available patients' questionnaires and completed CRFs per July 2014

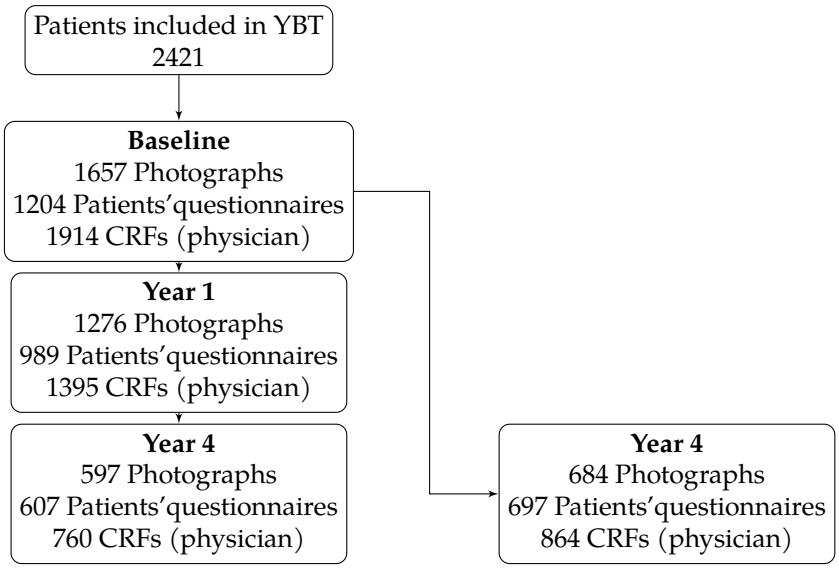


Table 4.1: Agreement between fibrosis scored by physicians and firmness scored by patients at four years, $w\kappa = 0.36$ (95% CI 0.29–0.42). Firmness was scored in the questionnaire of Sneeuw et al. (1992) by comparing the treated breast with the contralateral breast.

Physician's score of fibrosis	Patient's score of difference in firmness				
	No difference	A little	Quite a lot	Large	**Total**
No fibrosis	70	101	28	11	210
Mild fibrosis	70	157	63	19	309
Moderate fibrosis	19	96	77	25	217
Severe fibrosis	3	19	20	27	69
Total	162	373	188	82	805

scored by the patient. However, the presence of rib pain had no influence (table 4.3).

Of the EORTC QLQ C-30 questionnaire, we analysed whether emotional functioning, feelings of depression and/or global quality of life influenced PRCO. For the same BCCT.core score, patients with a higher emotional functioning or better global QoL had a better PRCO, whereas patients with feelings of depression had a worse PRCO (table 4.4).

Table 4.2: Proportional odds model for Patient Reported Cosmetic Outcome (PRCO) based on the seven dimensionless BCCT.core symmetry features described above (page 44). Significant p-values are indicated in bold. An odds ratio >1 means a worse PRCO.

Relative feature	Odds ratio	95% CI	p-value
pBRA	1.319	0.904–1.921	0.150
pLBC	1.142	0.792–1.648	0.477
pUNR	1.056	0.723–1.544	0.779
pBCE	1.177	1.008–1.375	**0.040**
pBCD	1.718	1.024–2.894	**0.041**
pBAD	0.856	0.540–1.352	0.505
pBOD	1.038	0.764–1.409	0.812

4.4 Discussion

The most important parameters related to PRCO after BCT in the YBT were the distance from nipple to inframammary fold and the length of breast contour. Also, the severity of fibrosis (physician) and the difference in firmness (patient) was related to the PRCO, independent of the BCCT.core score, suggesting that indeed a palpable firmness subjectively influences the patient's opinion on cosmesis. Rib pain was not related to the PRCO.

Comparison with the literature

Christie et al. (1996) found, in a population of 47 patients, that a greater degree of upward retraction of the nipple was the most powerful determinant of PRCO. This may be inversely related to the pBCE, i.e. the distance from nipple to inframammary fold, which we found in our study.

Patient's mental state might influence PRCO as well. Brunault et al. (2013) showed that depression is associated with patient-perceived cosmetic changes. Patients with a probable depression perceived the treated breast to be larger, more deformed and having worse skin pigmentation than non-depressed patients.

The current study finds similar results: patients with feelings of depression had worse PRCO than patients with better emotional functioning or better global QoL. However, it is difficult to distinguish between cause and effect in this matter. It might also be true that a better cosmetic result yields a better QoL. Recently the cosmetic results of the START trials were published. In this study, PRCO of 1870 patients was unaffected by anxiety and depression (Haviland et al., 2016). A possible explanation for the different findings could be the difference in age, since in the Cambridge IMRT trial young age was also found to influence the symptoms of

Table 4.3: Proportional odds model for patients' satisfaction with A: fibrosis scored by physician and BCCT.core score as covariate, B: difference in firmness scored by the patient and BCCT.core score as covariate and in C: rib pain and BCCT.core score as covariate. Odds ratios higher than 1 indicate that a higher value of the parameter was associated with a worse patient satisfaction.

	Odds ratio	95% CI	p-value
A			
BCCT.core score: 2	1.668	1.058–2.641	0.028
BCCT.core score: 3	3.856	2.348–6.372	<0.001
BCCT.core score: 4	9.479	4.835–18.73	<0.001
Physician fibrosis score: minor	1.183	0.797–1.760	0.404
Physician fibrosis score: moderate	2.022	1.314–3.121	0.001
Physician fibrosis score: severe	2.519	1.372–4.635	0.003
B			
BCCT.core score: 2	1.683	1.086–2.618	0.020
BCCT.core score: 3	2.735	1.695–4.431	<0.001
BCCT.core score: 4	4.616	2.427–8.812	<0.001
Patient difference firmness: small	1.700	1.152–2.516	0.008
Patient difference firmness: quite a lot	5.207	3.291–8.288	<0.001
Patient difference firmness: large	16.262	8.839–30.24	<0.001
C			
BCCT.core score: 2	1.772	1.140–2.765	0.011
BCCT.core score: 3	4.696	2.926–7.585	<0.001
BCCT.core score: 4	11.763	6.265–22.27	<0.001
Rib pain score: some	1.123	0.746–1.690	0.577
Rib pain score: moderate	0.945	0.235–2.690	0.915
Rib pain score: severe	1.988	0.531–7.567	0.306

skin appearance and breast hardness (Mukesh et al., 2014).

Correlation between BCCT.core and physician's opinion and/or patient's opinion

BCCT.core versus physician Cardoso et al. (2007) evaluated the validity of the BCCT.core software by a panel of experts. Overall inter-observer agreement for the subjective score was fair to moderate ($\kappa = 0.40$, w$\kappa = 0.57$), whereas the concordance level for the objective BCCT.core measurement was much higher ($\kappa = 0.86$, w$\kappa = 0.90$). The agreement between the subjective measurement and the BCCT.core was only fair ($\kappa = 0.34$, w$\kappa = 0.53$), but increased to moderate if scale 2 and 3 of the Harris scale were merged to a 3-point scale ($\kappa = 0.57$, w$\kappa = 0.72$). We found on a two-point scale, i.e. satisfactory or non-satisfactory overall cosmetic

Table 4.4: Proportional odds model for Patient Reported Cosmetic Outcome (PRCO) and quality of life with BCCT.core as covariate. A: emotional functioning with BCCT.core as covariate, B: feelings of depressing and C: global quality of life with BCCT.core as covariate. For BCCT core parameters, an odds ratio > 1 means a worse PRCO. For Emotional functioning, Global quality of life, and depression an odds ratio > 1 means worse PRCO in case of worse Emotional functioning, QoL, or more feelings of depression

	Odds ratio	95% CI	p-value
A			
BCCT.core score: 2	1.614	0.996–2.623	0.053
BCCT.core score: 3	4.885	2.902–8.286	<0.001
BCCT.core score: 4	8.507	4.283–17.05	<0.001
Emotional functioning	0.881	0.813–0.955	0.002
B			
BCCT.core score: 2	1.648	1.018–2.679	0.043
BCCT.core score: 3	4.825	2.867–8.179	<0.001
BCCT.core score: 4	9.250	4.670–18.50	<0.001
Feelings of depression	1.366	1.081–1.724	0.009
C			
BCCT.core score: 2	1.621	0.999–2.638	0.051
BCCT.core score: 3	4.709	2.796–7.988	<0.001
BCCT.core score: 4	8.618	4.318–17.35	<0.001
Global quality of life	0.790	0.685–0.909	0.001

outcome, a somewhat lower correlation between BCCT.core score and physician scores: 0.39. A possible explanation is that in the YBT the cosmetic evaluation was scored only by the treating physician instead of by a panel that reached consensus.

BCCT.core versus patient The correlation between objective measures and PRCO shows reported κ values varying from 0.04 to 0.34 (Heil et al., 2011; Yu et al., 2016), which corresponds to the value of 0.26 found in our study. The different κ values in the different studies can probably be ascribed to different methods to measure PRCOs. Yu et al. (2016) used a conversation with researchers not involved in treatment of patients. Heil et al. (2011) used a validated patient questionnaire BCTOS (Breast Cancer Treatment Outcome Scale), but another one than ours (Sneeuw et al., 1992). In both questionnaires (BCTOS and ours), patients were asked to rate seven items according to symmetry. In the BCTOS the rounded mean of these seven items was used as an overall score, while we compared only the last question in our questionnaire with the BCCT.core score, since that question dealt with the overall PRCO. The fact that the PRCO correlated less with the overall BCCT.core score than the physician's opinion, confirms our hypothesis that specific symmetry pa-

rameters were more important than others. Furthermore, we found that also other factors such as fibrosis, not directly measured by BCCT.core, influenced PRCO.

Correlation between physicians' and patients' opinion Several studies comparing patient with physician's reported overall cosmetic outcome showed various results. In some studies, similar to our study, patients scored their cosmesis and/or normal tissue effects worse than the clinician (Mukesh et al., 2014; Haviland et al., 2012) or photographic assessment(Haviland et al., 2012), while other studies, showed opposite results (Haloua et al., 2014; Hau et al., 2012).

The START trial (Haviland et al., 2016) also reported on agreement between PRCOs and clinical or photographic assessments of breast specific normal tissue effects. They found wκ coefficients of 0.05–0.21. These lower values might be explained by the difference in questions. For example, in the START trial telangiectasia (clinicians) was correlated with skin changes (patient), which could mean more than only telangiectasia (w$\kappa = 0.08$ at 5 years). Also, in some questions the patient was asked to indicate whether their scoring was influenced by radiotherapy, which is difficult if not impossible to judge by the patient. Another difference with our study was that the photographic assessment was performed by a panel, whilst we used an objective software program to analyse the photographs.

Overall cosmetic scores The overall cosmetic outcome in the YBT was worse than published in most other studies. Only Haloua et al. (2014) found similar results as we did. However, no data on radiation dose were given in this paper.

In the boost versus no boost trial 86% of the patients had excellent or good score in the no boost group compared to 71% in the boost group at 3 years (Vrieling et al., 1999), whereas in the YBT these scores were only found in 56–61%, dependent on the scoring method.

Better cosmetic outcome results are also reported by Kelemen et al. (2012) and Haloua et al. (2014), who found excellent/good cosmesis, scored by physician or patient in 95% and 93% in the boost- and no-boost-arm respectively) versus 81% (boost) and 68% (no boost) according to the BCCT.core software. In this trial the whole breast dose was lower in the boost arm than in the no-boost arm.

A possible explanation for the worse cosmetic outcome results in the YBT is that half of the patient population received a high (26 Gy) boost. Detailed analysis of the effect of these treatment related factors on overall cosmetic outcome will be performed and presented in a separate paper.

Strengths and limitations

This study is the largest study reported up till now addressing the question which objective parameters are related to PRCO. In a large subset of patients, three kinds of cosmetic analyses were performed. A limitation of this study is that it comprises only a subset of the total number of patients included in the YBT. This may have several causes, like the relatively short median follow up of 51 months and the usual delay for sending in CRFs. For some patients not all digital photographs were available, or not usable due to quality or technical issues. Since we only analysed quantitative variables, we expect that the missing data did not significantly affect our overall results.

Another important aspect to take into account is that all patients were 50 years or younger. It is thus not clear whether the same correlation exists in elderly patients.

Conclusion

Patient reported cosmetic outcome is mostly related to the distance from the nipple to the inframammary fold, the length of the breast contour, and by the severity of fibrosis. Patients with higher emotional functioning or better QoL scored their cosmesis better.

Acknowledgments

The data management of the YBT was supported by a grant of the Dutch Cancer Society (CKTO 2003-13). The analysis of cosmetic outcome in this paper was supported by a Pink Ribbon grant (2011. WO04.C94). The French part of the trial was funded by the French Ministry of Health PHRC2009 and PHRC2012 grants.

Further, the authors thank J. Cardoso and M. Cardoso for the use of the BCCT.core software program, B. Hanbeukers for analysing the majority of digital photographs using the BCCT.core. software, and L. Pronk en R. Muusers for data retrieval.

Also, we acknowledge and thank Jerôme Lemonnier, project leader in France, for coordinating the Young Boost Trial in France and the following for their active participation: M. van Hezewijk, Leiden (NL); M.J.C. van der Sangen Eindhoven (NL); M.C. Stenfert Kroese, Deventer (NL); J.J.Jobsen, Enschede (NL); J.M. Immink, Delft (NL); M.E. Mast, Den Haag (NL); F.M. Gescher, Den Haag (NL); N. Bijker, Amsterdam (NL); J.W.M. Mens, Rotterdam (NL); W.G.J.M. Smit, Leeuwarden (NL); D.H.F. Rietveld, Amsterdam (NL); I. Lecouillard, Rennes (Fr); C Breton-Callu, Bordeaux (Fr); H. Marsiglia, Villejuif (Fr); J. Thariat, Nice (Fr); A. Benyoucef, Rouen (Fr); A. Labib, Saint Cloud (Fr); M. Aumont, Saint Herblain (Fr); P. Bontemps, Besancon (Fr); C. Le Foll, Lagny (Fr); Y. Belkacemi, Créteil (Fr); O. Chapet, Lyon (Fr); V. Strnad, Erlangen (De).

4.5 Supplementary Material

PRCO questionnaire The following questions have to do with the look and feel of your treated breast.

1. How would you rate the scar on your breast?
   1 = no noticeable scar
   2 = slightly noticeable scar
   3 = moderately noticeable scar
   4 = very noticeable scar
2. How would you compare the size of your treated breast with that of your untreated breast?
   1 = no difference
   2 = a small difference
   3 = a moderate difference
   4 = a large difference
3. How would you compare the shape of your treated breast with that of your untreated breast?
   1 = no difference
   2 = a small difference
   3 = a moderate difference
   4 = a large difference
4. Hw would you compare the firmness of your treated breast with that of your untreated breast?
   1 = no difference
   2 = a small difference
   3 = a moderate difference
   4 = a large difference
5. How would you compare the skin colour of your treated breast with that of your untreated breast?
   1 = no difference
   2 = a small difference
   3 = a moderate difference
   4 = a large difference
6. How would you compare the position of the nipple of your treated breast with that of your untreated breast?
   1 = no difference
   2 = a small difference
   3 = a moderate difference

4 = a large difference

7. How would you compare the overall appearance of your treated breast with that of your untreated breast?
   1 = no difference
   2 = a small difference
   3 = a moderate difference
   4 = a large difference
8. Overall, how satisfied or dissatisfied are you are you with the appearance of your treated breast?
   1 = very satisfied
   2 = satisfied
   3 = not dissatisfied
   4 = dissatisfied
   5 = very dissatisfied

CHAPTER 5

Predictors for Poor Cosmetic Outcome in Patients With Early-Stage Breast Cancer Treated With Breast Conserving Therapy

This chapter was published as:
Patricia J. A. M. Brouwers*, Erik van Werkhoven*, Harry Bartelink, Alain Fourquet, Claire Lemanski, Judith van Loon, John H. Maduro, Nicola S. Russell, Luc J. E. E. Scheijmans, Dominic A. X. Schinagl, Antonia H. Westenberg, Philip Poortmans†, and Liesbeth J. Boersma† on behalf of the Young Boost Trial research group.
Predictors for Poor Cosmetic Outcome in Patients With Early Stage Breast Cancer Treated With Breast Conserving Therapy: Results of the Young Boost Trial
Radiotherapy and Oncology, 128(3):434–411, 2018.
*) Equally contributing

Abstract

Purpose In the Young Boost trial (YBT), breast cancer patients ≤ 50 years of age, treated with breast conserving therapy (BCT) were randomized between a 26 Gy boost dose and a 16 Gy boost dose, with local recurrence as primary and cosmetic outcome (CO) as secondary endpoint. Data of the YBT was used to investigate which factors are related with worse cosmetic outcome after BCT.

Methods From 2004 to 2011, 2421 cT1-2N0-2a breast-cancer patients were randomized. CO was scored subjectively by the patient and physician, and objectively using BCCT.core: at baseline, one, and four years after treatment. Associations between potential risk factors for worse cosmetic outcome, based on the objective BCCT.core, were investigated using a proportional odds model.

Results At four years, CO was significantly better in the standard boost group for all three scoring methods (satisfied CO $\sim$65% vs 55%). A photon boost, high boost dose, poor cosmesis before radiation therapy, large boost volume and adjuvant chemotherapy significantly deteriorated CO.

Conclusion Important risk factors for worse CO were the use of a photon boost instead of an electron boost, a high boost dose, cosmesis at baseline, adjuvant chemotherapy and boost volume. These results can be used to define strategies aimed at improving CO.

5.1 Introduction

In women with early breast cancer treated with breast-conserving surgery (BCS), whole breast radiation therapy (RT) reduces the risk of local recurrence at 5 years from 26% to 7% (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005). The EORTC 'boost versus no boost' trial showed that an additional boost of 16 Gy to the tumour bed reduces the risk for local failure by a factor of 2, with an increased incidence of moderate/severe fibrosis as negative side effect (Poortmans et al., 2008). However, after 10 years of follow-up, the risk of local failure remained unacceptably high, in the younger patients, even after a boost, with a risk of 13.5% in patients ≤ 40 years, and of 8.7% in patients 41–50 years (Bartelink et al., 2007).

Therefore, in 2004, the Young Boost trial (YBT) was launched (NCT00212121) with the primary aim to investigate whether a higher boost dose of 26 Gy to the tumour bed would further reduce local recurrence rate in these young patients with cosmetic outcome as secondary endpoint.

Several risk factors for deterioration of the cosmetic outcome have been described in literature, for example breast size (Barnett et al., 2011; Peterson et al., 2015), tumour size and excision volume (Vrieling et al., 1999; Immink et al., 2012), tumour location (Peterson et al., 2015; Vrieling et al., 1999; Immink et al., 2012), post-operative complications (Barnett et al., 2011; Peterson et al., 2015), boost volume (Mukesh et al., 2013a), a photon boost (Immink et al., 2012; Collette et al., 2008), total dose (Vrieling et al., 1999) and dose max (Mukesh et al., 2013a; Collette et al., 2008; Hammer et al., 2017). However, no data are available concerning a boost dose as high as 76 Gy EQD2, which makes the YBT unique. Moreover, in order to be able to improve cosmetic outcome, we need to continue to update the knowledge of risk factors for cosmetic outcome with data derived from the most current literature.

It was decided by the independent data monitoring committee that the primary endpoint (i.e. local failure) should not be analysed yet. However, they recommended that the cosmetic outcome, which was a secondary endpoint, could be analysed by treatment arm now that up to 4 years of follow-up is available. Previously, we reported that the distance from nipple to inframammary fold, the length of the breast contour and the severity of fibrosis were associated with patient reported outcome in the YBT (Brouwers et al., 2016). The primary aim of this paper is to report on the cosmetic outcome in the YBT; the secondary aim is to define risk factors for worse cosmetic outcome in this patient population, based on the objective BCCT.core.

5.2 Patients and Methods

Patient population and treatment

Patients younger than 51 years with non-metastatic, histological proven invasive breast cancer, pT1-2N0-2a (Sobin et al., 2002) were eligible for the trial if fulfilling the following inclusion criteria: ECOG performance scale ≤ 2; wide local excision (WLE); microscopically complete (no tumour on ink) or focally involved (defined as: 'tumour (ductal carcinoma in situ or invasive carcinoma) on ink in an area of less than 4 mm') resection; sentinel lymph node biopsy and/or axillary lymph node dissection; no primary systemic treatment; no previous history of malignant disease, except adequately treated carcinoma in situ of the cervix or basal cell carcinoma of the skin. Exclusion criteria were: residual microcalcifications on mammogram; histological other than invasive adenocarcinoma; in situ carcinoma of the breast without invasive tumour; multicentric tumours and multifocal tumours excised using multiple excisions; bilateral invasive breast cancer and concurrent pregnancy. More information can be found at `https://clinicaltrials.gov/show/NCT00212121`.

Patients were randomized to receive a standard 16 Gy or a high 26 Gy boost to the tumour bed after 50 Gy whole breast irradiation, given in 2 Gy fractions. Other fractionation schemes, including simultaneous integrated boost (SIB) techniques were allowed as well, as long as the biological equivalent dose (EQD2), calculated with an α/β of 10 for tumour control, was similar. The overall treatment time was kept constant in both randomization arms, i.e. 6.5–7 weeks (see page 75 for more extensive information concerning the RT protocol). RT had to start within 10 weeks after surgery. In case adjuvant chemotherapy was given immediately after surgery, RT should start within 6 months after surgery and within 6 weeks after the last cycle of chemotherapy. In case endocrine treatment was planned, this was recommended to start after completion of the RT. Stratification factors were age ($\leq$ vs. $>$ 40 yrs.), pathological tumour size ($\leq$ vs. $>$ 3 cm), oestrogen receptor status, nodal status, interstitial/external boost and institute. Patients were stratified at the time of randomization; treatment was assigned using a minimization technique (Scott et al., 2002). The study was centrally approved by the medical ethical committee of the Netherlands Cancer Institute and by the local medical ethics committees. All patients gave their written informed consent to participate.

Recording of fibrosis and cosmetic outcome

Cosmetic outcome and fibrosis were scored at baseline, i.e. after surgery but prior to start of RT, at 1 year, 4, 7, and 10 years of follow-up (FU). Standardized digital photographs were taken at the same time points.

The presence of fibrosis (whole breast and specifically in the boost area) was scored by the physician on a 4-point scale: none, minor, moderate or severe.

Cosmetic outcome was scored according to the following three scoring systems:

- *BCCT.core software* (Cardoso et al., 2007; Cardoso & Cardoso, 2007): digital photographs in anterior-posterior view were analysed using the BCCT.core software program. Predetermined points were designated by the examiner, followed by an automatic calculation of an overall cosmetic score: excellent, good, fair or poor (score 1–4; higher score means worse outcome). This score is based on symmetry, skin colour and scar visibility.
- *Physician's score*: Physicians scored cosmetic outcome using the Harris scale (Harris et al., 1979): excellent, good, fair or poor, indicated as score 1–4 respectively.
- *Patient's questionnaire*: Patients' satisfaction with the cosmetic outcome was scored using a validated patient's questionnaire developed by Sneeuw et al. (1992): very satisfied, satisfied, not dissatisfied, dissatisfied or very dissatisfied (score 1–5 respectively). For the analyses of crude percentages, the scores very satisfied or satisfied and good or excellent were grouped as 'satisfactory'.

Analysis of risk factors for fibrosis and cosmetic outcome

The following risk factors, scored on the Case Report Forms, were investigated:

- *RT related risk factors:* dose to the tumour bed; irradiated boost volume (per 10 cc), defined as the volume receiving more than 95% of the boost dose for external photon irradiation, and within 85% of the boost dose for electron and interstitial irradiation; photon boost versus electron boost; Simultaneous Integrated Boost (SIB) versus sequential boost; energy used for whole breast irradiation (WBI) and the use of CT-scan for planning.
- *Systemic therapy related factors:* adjuvant chemotherapy, adjuvant endocrine therapy.
- *Surgery related factors:* excision volume (per 10 cc); post-operative complications and seroma, scored as yes, no, or unknown. Postoperative complications were defined as the presence of infection and/or haematoma of breast and/or axilla. Oedema was not considered as a complication. Seroma was analysed separately from post-operative complications, as we assumed there might be a correlation with oncoplastic surgery.
- *Tumour related factors:* tumour location (lateral tumour location vs. central and medial/upper tumour location vs. central) (figure 5.3 on page 74).
- *Patient characteristics:* age (per year) and cosmetic score at baseline.

Statistical analysis

The percentages of patients with satisfactory cosmetic scores in the high- and standard-boost group were compared at baseline, 1 year, and 4 years with Fisher's exact test. Associations between potential risk factors and cosmetic outcome, measured by BCCT.core, were assessed with a proportional odds model, in order to treat the cosmetic outcome as a variable with ordered categories. An important assumption of the proportional odds model is that the association between each pair of outcome groups is the same, so that for example the comparison between a score of 1 (=Excellent) versus a score of 2 (=Good), 3 (=Fair) or 4 (=Poor), and the comparison of 1 or 2 versus 3 or 4 can be modeled by the same parameter. The assumption was verified by calculation of linear predictions from a logit model, used to model the probability that the outcome is greater than or equal to a given value (for each cosmetic outcome level). These were compared between categories of one predictor variable at a time, and no great differences were observed.

Both the number of patients with moderate and severe fibrosis, and of patients with severe fibrosis at baseline, 1 year, and 4 years was calculated as a percentage of the total number of patients with an assessment and compared by arm using Fisher's test. Time to fibrosis was calculated from randomization to first reported occurrence of moderate or severe fibrosis. Patients with no or only minor fibrosis were censored at last follow-up. Risk factors for moderate or severe fibrosis were analysed with multivariable Cox proportional-hazards models.

5.3 Results

Between 2004 and 2011, 2421 breast cancer patients were included in 32 institutes (18 from The Netherlands, 13 from France and 1 from Germany). 1211 patients were randomized to receive a standard 16 Gy boost and 1210 to receive a 26 Gy boost. Baseline patient characteristics were similar in both groups with the exception of boost technique (table 5.1). Median age was 45 years (range 19–51), 19% was younger than 40 years of age. 72% of patients had a T1 tumour and 28% of patients had a T2 tumour. Median follow-up at the time of this analysis was 51 months. 46 patients did not comply with the inclusion criteria (table 5.6 on page 74). All patients with available and evaluable digital photographs were included in the analysis.

Table 5.1: Patient and treatment characteristics at baseline

	Randomized treatment			
	16 Gy boost 1211	26 Gy boost 1210	Total 2421	p-value
Age				0.94[1]
median (range)	45 (19–51)	45 (21–51)	45 (19–51)	
Age (grouped)				0.99[2]
[19,25)	1 (0.1%)	2 (0.2%)	3 (0.1%)	
[25,30)	15 (1.2%)	13 (1.1%)	28 (1.2%)	
[30,40)	219 (18.1%)	223 (18.4%)	442 (18.3%)	
[40,45)	348 (28.7%)	351 (29.0%)	699 (28.9%)	
[45,50)	516 (42.6%)	512 (42.3%)	1028 (42.5%)	
[50,51]	112 (9.2%)	109 (9.0%)	221 (9.1%)	
Tumour location				0.69[3]
central-under	275 (22.8%)	293 (24.3%)	568 (23.6%)	
lateral	606 (50.3%)	594 (49.3%)	1200 (49.8%)	
medial-up	323 (26.8%)	317 (26.3%)	640 (26.6%)	
NA	7	6	13	
Pathological largest diameter (mm)				0.73[1]
median (range)	(1–49)	15 (1–95)	15 (1–95)	
NA	6	9	15	
Largest diameter (grouped)				0.47[3]
$\leq$ 20 mm	860 (71.4%)	874 (72.8%)	1734 (72.1%)	
$>$ 20 mm	345 (28.6%)	327 (27.2%)	672 (27.9%)	
NA	6	9	15	
Excision volume (ml)				0.19[1]
median (range)	112 (0.06–3150)	105 (0.16–4462)	108 (0.06–4462)	
NA	110	90	200	
Final margin status				0.79[3]
Complete	1180 (97.4%)	1182 (97.7%)	2362 (97.6%)	
Focally incomplete	31 (2.6%)	28 (2.3%)	59 (2.4%)	
Extensive involvement	0 (0.0%)	0 (0.0%)	0 (0.0%)	
Postoperative complications				0.29[3]
No	818 (70.9%)	835 (73.0%)	1653 (72.0%)	
Yes	335 (29.1%)	309 (27.0%)	644 (28.0%)	
NA	58	66	124	
Endocrine therapy*				0.64[3]
No	483 (42.0%)	491 (43.0%)	974 (42.5%)	
Yes	667 (58.0%)	650 (57.0%)	1317 (57.5%)	
NA	61	69	130	
Chemotherapy				0.37[3]
No	441 (37.1%)	458 (38.9%)	899 (38.0%)	
Yes	748 (62.9%)	719 (61.1%)	1467 (62.0%)	
NA	22	33	55	

(Continued on next page)

Table 5.1 – continued from previous page

	16 Gy boost 1211	26 Gy boost 1210	Total 2421	*p*-value
Chemotherapy timing				0.08[3]
Prior to RT	364 (48.9%)	357 (49.9%)	721 (49.4%)	
During RT	10 (1.3%)	2 (0.3%)	12 (0.8%)	
After RT	370 (49.7%)	356 (49.7%)	726 (49.7%)	
Other	0 (0.0%)	1 (0.1%)	1 (0.1%)	
NA	467	494	961	
WBI quality				0.12[3]
Cobalt60	0 (0.0%)	3 (0.3%)	3 (0.1%)	
X-ray beams	1196 (100.0%)	1180 (99.7%)	2376 (99.9%)	
NA	15	27	42	
X-ray energy (MV) WBI				0.951[3]
median (range)	6 (4–25)	6 (4–25)	6 (4–25)	
NA	125	155	280	
Irradiated boost volume (cc)				0.080[1]
median (range)	135 (0–1125)	130 (0-1308)	132 (0–1308)	
Boost technique				0.04[3]
Electrons	265 (22.1%)	214 (18.2%)	479 (20.2%)	
Cobalt60	6 (0.5%)	4 (0.3%)	10 (0.4%)	
X-Ray beams	882 (73.7%)	895 (76.0%)	1777 (74.8%)	
Interstitial boost	10 (0.8%)	13 (1.1%)	23 (1.0%)	
Other	34 (2.8%)	52 (4.4%)	86 (3.6%)	
NA	14	32	46	
SIB				0.83[3]
No	784 (65.3%)	768 (64.9%)	1552 (65.1%)	
Yes	416 (34.7%)	416 (35.1%)	832 (34.9%)	
NA	11	26	37	
Planning CT				0.74[3]
No	286 (23.8%)	291 (24.4%)	577 (24.1%)	
Yes	917 (76.2%)	902 (75.6%)	1819 (75.9%)	
NA	8	17	25	

*) in 85% Tamoxifen
[1]) Kruskal-Wallis rank sum test
[2]) Pearson's Chi-squared test
[3]) Fisher's Exact Test for Count Data

At baseline, 1657 evaluable digital photographs were available of the study population. At one year, evaluable digital photographs were available from 1455 patients, of whom 1276 also had an evaluable photograph at baseline. At four years, 684 digital photographs were evaluable of patients including a photograph at baseline (figure 5.1).

At baseline, cosmetic score was similar in both patient groups independent of the scoring methods. In 90% cosmetic score was satisfactory based on BCCT.core. According to the physician or patient, satisfactory scores were 80% or a little less than 70% respectively, at baseline. At 4 years, cosmetic outcome was significantly worse

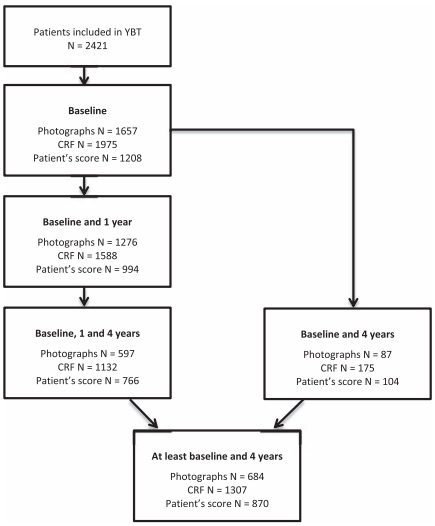


Figure 5.1: Flow diagram of available and evaluable digital photographs per July 2014, and completed Case Report Form (CRF) and completed patient questionnaires of all institutes per February 2017

than at baseline in both treatment arms, for all three scoring methods. The cosmetic outcome was better in the standard boost group compared to the high boost group for all three scoring methods: according to BCCT.core 67% of patients had satisfactory cosmesis in the standard boost, versus 55% in the high boost group (p=0.0009). For scores by the physicians these numbers were 65% and 52% ($p < 0.0001$), and for patients 63% and 53% ($p = 0.0007$), respectively (table 5.2a).

At 4 years, the physician scored moderate or severe fibrosis in the boost area in 159 patients (19%) in the standard boost group versus 332 (39%) in the high boost group ($p < 0.0001$). Severe fibrosis was scored in the boost area in 25 (3%) and 89 (11%) patients in the standard and high boost group, respectively ($p < 0.0001$, table 5.2b). Also, when fibrosis was calculated as a percentage of the evaluable patients at the three time points separately, the difference between the arms remained significant (table 5.2c).

The cumulative incidence of moderate or severe fibrosis in the boost area at 4 years was 27% (95%CI 24–30%) in the low boost group versus 45% (95%CI 42–47%) in the high boost group ($p < 0.0001$, figure 5.2).

Significant risk factors in the multivariable model for worse cosmetic outcome according to BCCT.core score at 4 years were a photon boost (odds ratio 1.98 compared to electrons), a high boost dose (odds ratio 1.82 compared to standard boost),

Table 5.2: Outcome at baseline, one year, and four years of follow-up w/ numbers of patients as percentage of patients with available score (%)

Satisfactory cosmesis			
	Baseline		
	16 Gy	26 Gy	p-value
BCCT.core	741/831 (89%)	745/826 (90%)	0.52
Physician	774/970 (80%)	771/988 (78%)	0.35
Patient	415/604 (69%)	406/604 (67%)	0.62
	1 year		
	16 Gy	26 Gy	p-value
BCCT.core	490/702 (70%)	442/706 (63%)	0.0048
Physician	616/906 (68%)	559/941 (59%)	0.00013
Patient	441/666 (66%)	410/674 (61%)	0.0007
	4 years		
	16 Gy	26 Gy	p-value
BCCT.core	265/397 (67%)	225/408 (55%)	0.0009
Physician	484/749 (65%)	391/753 (52%)	< 0.0001
Patient	361/577 (63%)	307/584 (53%)	0.0007

(**a**) Satisfactory Cosmesis

cosmesis at baseline (odds ratio 1.80 per BCCT.core category), adjuvant chemotherapy (odds ratio 1.58 yes vs. no) and boost volume (odds ratio 1.04 per 10 cc). The following factors were not significantly associated with cosmetic outcome: age, tumour location, adjuvant endocrine therapy, radiation energy WBI, use of CT for planning, excision volume per 10 cc, postoperative complications, seroma or whether the boost was given simultaneously (SIB) versus sequentially (table 5.3).

Significant risk factors for moderate or severe fibrosis were cosmesis at baseline (HR 1.20 per BCCT.core category), a high boost dose (HR 2.00), age (HR 1.02 per year older), adjuvant chemotherapy (HR 1.25 yes vs. no), radiation energy WBI (HR 1.03 per MV), irradiated boost volume (HR 1.01 per 10 cc) and a simultaneous integrated boost (HR 1.40 yes vs. no, table 5.4).

Table 5.2: (continued)

Fibrosis (whole breast)			
	Baseline		
	16 Gy	26 Gy	p-value
None or minor	817/834 (98%)	839/850 (99%)	0.26*
Moderate or severe	17/834 (2%)	11/850 (1%)	0.26*
Severe	1/834 (0%)	2/850 (0%)	1.00
	1 year		
	16 Gy	26 Gy	p-value
None or minor	1007/1062 (95%)	951/1042 (91%)	0.0015*
Moderate or severe	55/1062 (5%)	91/1042 (9%)	0.0015*
Severe	8/1062 (1%)	8/1042 (1%)	1.00
	4 years		
	16 Gy	26 Gy	p-value
None or minor	829/854 (97%)	787/848 (93%)	< 0.0001*
Moderate or severe	25/854 (3%)	61/848 (7%)	< 0.0001*
Severe	3/854 (0%)	9/848 (1%)	0.09

*) testing none/minor vs. moderate/severe

(**b**) Fibrosis in the whole breast

5.4 Discussion

The results of this analysis demonstrate that, as expected, a high boost causes a less satisfactory cosmetic outcome. At 4 years of follow-up, the percentage of patients with a satisfactory cosmetic outcome was about 10% lower in the high boost group compared to the standard boost group, whichever scoring method (BCCT.core, physician, or patient herself) was used. Also, in the high boost group twice as much moderate or severe fibrosis was scored at 4 years. The multivariable model showed that other important risk factors for worse cosmetic outcome were the use of a photon boost, cosmesis at baseline, adjuvant chemotherapy and boost volume. It is important to note that we have reported the estimate of the effect of the boost volume as a continuous variable per 10 cc. This means that the odds ratio holds for every increase of 10 cc. The odds ratio is 1.48 if the boost volume is considered per 100 cc.

Table 5.2: (continued)

Fibrosis (boost area)			
	Baseline		
	16 Gy	26 Gy	p-value
None or minor	777/826 (94%)	782/851 (92%)	0.09
Moderate or severe	49/826 (6%)	69/851 (8%)	0.09
Severe	3/826 (0%)	3/851 (0%)	1.00
	1 year		
	16 Gy	26 Gy	p-value
None or minor	898/1053 (85%)	783/1038 (75%)	<0.0001*
Moderate or severe	155/1053 (15%)	255/1038 (25%)	<0.0001*
Severe	21/1053 (2%)	38/1038 (4%)	0.024
	4 years		
	16 Gy	26 Gy	p-value
None or minor	690/849 (81%)	509/841 (61%)	<0.0001*
Moderate or severe	159/849 (19%)	332/841 (39%)	<0.0001*
Severe	25/849 (3%)	89/841 (11%)	<0.0001

*) testing none/minor vs. moderate/severe

(**c**) Fibrosis in the boost area

Risk factors for moderate or severe fibrosis consisted of the same risk factors as for worse cosmetic outcome with the exception of a photon boost and supplemented with age, photon energy of WBI en a simultaneous integrated boost technique.

Although a worse cosmetic outcome was expected for the high boost arm, we surprisingly also observed a somewhat worse cosmetic outcome in the standard boost arm, compared to the identical 16 Gy boost arm in the former boost versus no-boost trial (Vrieling et al., 1999). In the latter trial, the panel evaluation at 3 years showed that 71% of patients in the boost group had an excellent or good global result, which is better than the 65% satisfactory score by the physicians that we found in the 16 Gy boost arm. There are several possible explanations for this difference. First, in the majority of cases in the boost versus no-boost trial, the boost dose was given with electrons (74.9%) (Immink et al., 2012); whereas a photon boost was the most important risk factor in our model. Second, in the YBT only the treating physician scored the cosmesis, in contrast to the boost vs no-boost trial, where cosmetic

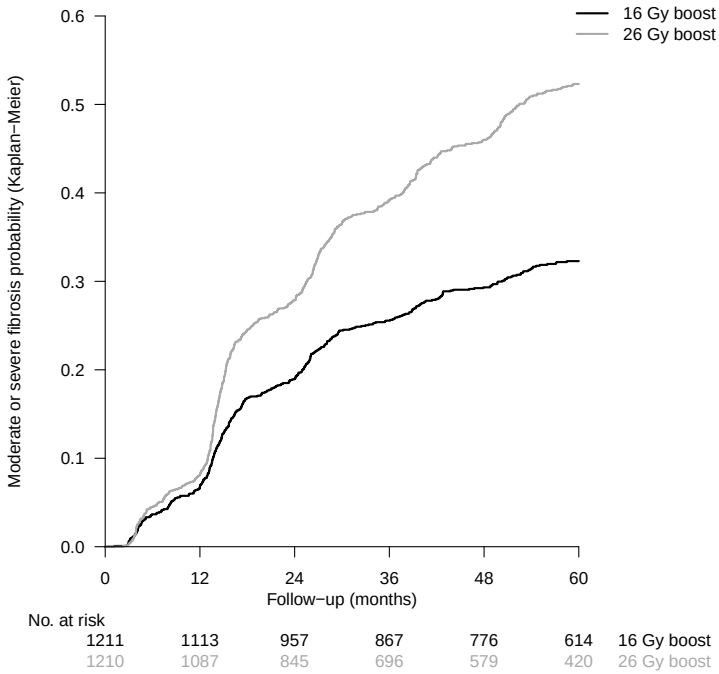


Figure 5.2: Cumulative incidence of moderate or severe fibrosis in the boost area

outcome was scored by a panel. Third, in the YBT timing of scoring was one year later (at four instead of three years follow-up); the boost vs. no-boost trial already showed that asymmetry progressed over years (Immink et al., 2012). Fourth, in the YBT, a larger amount of patients underwent chemotherapy. In the boost no-boost only 10% of patients received chemotherapy, in the YBT this percentage was 60% and chemotherapy was identified as a risk factor for worse cosmetic outcome in our model. Finally, also the boost volume was different. Al Uwini et al. (2009) already showed an enlargement of boost volumes using a planning CT. He recalculated the boost volumes of the boost versus no-boost trial and showed that the volume of the 95% dose level was larger in the YBT. Surprisingly, use of a planning CT was not an independent risk factor in our model, but there might be interaction with the volume variable.

Previous studies found various risk factors for worse cosmetic outcome or fibrosis. Cosmetic outcome and fibrosis are both late toxicity endpoints and are probably associated with each other, but show different progression in time. Where fibro-

Table 5.3: Multivariable proportional odds model for cosmetic outcome based on BCCT.core. Odds ratio > 1 means a negative impact on cosmetic outcome, < 1 a positive impact

	Odds ratio	95% CI	p-value
Cosmesis at baseline	1.80	1.40–2.33	**<0.0001**
High boost dose	1.83	1.33–2.54	**<0.0001**
Age (per year)	0.99	0.96–1.02	0.557
Tumour location			
Lateral vs. central	0.70	0.47–1.03	0.073
Medial/upper vs. central	0.83	0.53–1.31	0.429
Adjuvant chemotherapy	1.53	1.04–2.27	**0.032**
Adjuvant endocrine therapy	1.16	0.80–1.69	0.429
Photon energy of WBI	1.05	0.97–1.13	0.232
Boost volume per 10 cc*	1.04	1.02–1.05	**<0.0001**
Boost technique			
photon vs. electron	1.98	1.31–3.01	**<0.0001**
SIB vs. sequential boost	0.96	0.63–1.46	0.837
Seroma	1.52	0.93–2.50	0.097
Postoperative complications	1.15	0.78–1.70	0.478
Excision volume per 10 cc	1.00	1.00–1.01	0.448
Use of planning CT	0.90	0.62–1.31	0.585

WBI = whole breast irradiation; SIB = simultaneous integrated boost

*) The irradiated boost volume was defined as the volume receiving more than 95% of the boost dose for external photon irradiation and within 85% of the boost dose for electron and interstitial irradiation

sis is most progressive in the first three to four years (Collette et al., 2008), cosmetic deterioration progresses further over the years, also resulting from increasing asymmetry following more pronounced changes in the non-treated breast with ageing (Immink et al., 2012). The results in literature are difficult to interpret due to different outcome measures including fibrosis and cosmetic outcome (automatic photograph based, patient score, panel or physician score) and different duration of follow-up. Nevertheless, all various risk factors can be brought together to some overarching risk factors: 1. Dose homogeneity (IMRT (Mukesh et al., 2013b), Dmax (Mukesh et al., 2013a; Collette et al., 2008; Hammer et al., 2017), V55Gy (Hammer et al., 2017), V110 (Lazzari et al., 2017), V107 (Barnett et al., 2011), breast size (Barnett et al., 2011; Peterson et al., 2015), prone/supine (Veldeman et al., 2016)); 2. Total dose (hypofractionation (Haviland et al., 2013; Whelan et al., 2010), boost no-boost (Vrieling et al., 1999), Young Boost); 3. Boost volume (Mukesh et al., 2013a) (excision volume (Vrieling et al., 1999; Immink et al., 2012), tumour size (Vrieling et al., 1999; Immink et al., 2012), photon boost (Immink et al., 2012; Col-

Table 5.4: Multivariable model of time to moderate or severe fibrosis in the boost area

	HR	95% CI	p-value
Cosmesis at baseline	1.20	1.06–1.35	**0.003**
High boost dose	2.00	1.71–2.35	**<0.0001**
Age at randomization	1.02	1.01–1.04	**0.005**
Tumour location			
Lateral vs. central	0.98	0.80–1.19	0.081
Medial/upper vs. central	1.16	0.94–1.44	0.17
Adjuvant chemotherapy	1.25	1.04–1.51	**0.017**
Adjuvant endocrine therapy	0.97	0.81–1.15	0.72
Photon energy of WBI	1.03	1.01–1.06	**0.007**
Boost volume per 10 cc*	1.01	1.01–1.02	**<0.0001**
Boost technique			
(photon vs. electron)	1.13	0.90–1.40	0.30
SIB vs. sequential boost	1.40	1.16–1.71	**0.0006**
Seroma	1.19	0.96–1.47	0.11
Postoperative complications	1.05	0.87–1.27	0.62
Excision volume per 10 cc	1.00	1.00–1.00	0.28
Use of planning CT	0.89	0.73–1.10	0.28

WBI = whole breast irradiation; SIB = simultaneous integrated boost

*) The irradiated boost volume was defined as the volume receiving more than 95% of the boost dose for external photon irradiation and within 85% of the boost dose for electron and interstitial irradiation

lette et al., 2008), re-excision (Bantema-Joppe et al., 2012), time between surgery and RT, oncoplastic surgery (Lansu et al., 2015)) and 4. Baseline cosmesis (excision volume (Vrieling et al., 1999; Immink et al., 2012), tumour size (Vrieling et al., 1999), location of tumour (Peterson et al., 2015; Vrieling et al., 1999; Immink et al., 2012), post-operative complications (Barnett et al., 2011; Peterson et al., 2015)). Further, adjuvant chemotherapy might result in worse cosmesis (Vrieling et al., 1999; Bantema-Joppe et al., 2012; Keller et al., 2012). However, nowadays, many patients receive primary chemotherapy and one can assume this beneficially influences the cosmetic result by decreasing tumour size, resulting in smaller excision volumes (better baseline cosmesis).

We were somewhat surprised to find SIB as a risk factor for moderate or severe fibrosis, as several planning studies showed dosimetric advantage (Franco et al., 2015). To our knowledge, only the group of Groningen published data concerning fibrosis in a large cohort of breast cancer patients treated with a photon SIB (Hammer et al., 2017; Bantema-Joppe et al., 2012). They found moderate or severe

fibrosis in maximal 13.4% of patients, compared to the 22% (data not shown) we found in our standard boost arm (SIB), but they did not compare it with sequential boost results. One explanation might be that the fraction size to the boost volume was higher with the SIB than with the sequential boost, resulting in a higher EQD2 (67.6 vs 66 Gy, and 78.5 Gy vs 76 Gy for an α/β ratio of 4 Gy, and 68.2 Gy vs 66 Gy and 79.5 Gy vs 76 Gy vs for an α/β ratio of 3 Gy.

Unfortunately, we did not score whether oncoplastic surgery had been performed. The obvious aim of oncoplastic surgery is to improve cosmetic outcome. However, after oncoplastic breast surgery the definition of the tumour bed could be more difficult, because of large mammary gland translations, rotations or excisions. Therefore, tumour bed delineation after oncoplastic surgery will be difficult, especially without surgical clips (González Sanchis et al., 2013), which can lead to larger boost volumes (Furet et al., 2014). Close collaboration between surgeon and radiation oncologist could lead to a reliable, compact boost volume after oncoplastic surgery (mark lumpectomy cavity, then approximate lumpectomy cavity, then apply oncoplastic manoeuvres). The challenge for the future is to find an accurate balance between the extent of oncoplastic surgery and the following uncertainties for the radiation oncologist (Boersma et al., 2012).

It could have been interesting to analyse the impact of the timing of chemotherapy on cosmetic outcome. We tried to analyse this in the multivariable model by putting chemotherapy into the model as a variable with three categories: chemotherapy before RT, after RT and no chemotherapy at all. This showed that compared to no chemotherapy at all, chemotherapy before RT was significantly associated with worse cosmetic outcome, but not if the chemotherapy was given after RT (results not shown). In order to clarify this discrepancy, we looked within the subgroup of patients with chemotherapy. In that subgroup, there was no difference between before and after, whether we corrected for the other clinical variables in the model or not. Therefore, we believe that we do not have sufficient power to draw valid conclusions about the impact of the timing of chemotherapy.

The Young Boost Trial is a large international randomized trial and by our knowledge the only trial to investigate the influence of such a high boost dose (EQD2 76 Gy) on cosmetic outcome. Nevertheless, there are some limitations to mention. First of all, we were unable to test all the now known risk factors, such as for example smoking and breast size, since these factors were not known during the design of the YBT. Further, as we described in the methods section, the study was designed with an α/β of 10 for tumour control, which was a logical assumption at that time. However, the START trials have shown an α/β value for locoregional relapse of 3.5 Gy (Haviland et al., 2013). The results of the YBT provide better perception of the risk factors for worse cosmetic outcome. These data therefore provide valuable tools when developing a strategy to improve cosmetic outcome. Since boost dose

was one of the most important risk factors predicting poor cosmetic outcome, and local control has increased considerably in the last decade (Van der Heiden-van der Loo et al., 2015; Poortmans et al., 2017), we advise to critically re-evaluate the indication for a (high) boost. Whenever a boost is indicated, an electron boost might be preferred, on the condition that the boost volume is delineated (instead of virtual simulation). Further, the size of the boost volume should be limited as much as possible, using all available pre- and post-operative data (Boersma et al., 2012; Kirova et al., 2010). How to take into account baseline cosmetic score is however puzzling: one may argue that oncoplastic surgery will improve cosmetic outcome, since a good baseline cosmesis is correlated with a better cosmetic outcome; however, some studies also suggest that oncoplastic surgery leads to a worse cosmetic outcome (Lansu et al., 2015), possibly as a consequence of the resulting larger boost volumes combined with more tissue damage due to extended devascularization of the intramammary tissue flaps. The most important issues that need further studies are both the influence of extensive oncoplastic surgery and the influence of primary chemotherapy on cosmetic outcome.

In conclusion, the 4 year results of the YBT show that a photon boost, a high boost dose, poor cosmesis before RT, large boost volume, and adjuvant chemotherapy result in worse cosmetic outcome. These data offer valuable tools to develop strategies aimed at improving cosmetic outcome.

Acknowledgements

This publication was supported by a Pink Ribbon grant (2011. WO04.C94). The French part of the trial was funded by the French Ministry of Health PHRC2009 and PHRC2012 grants.

Further, the authors thank J. Cardoso and M. Cardoso for the use of the BCCT.core software program, and B. Hanbeukers for analysing the majority of digital photographs using the BCCT.core. software.

Also, we acknowledge and thank Jerôme Lemonnier, project leader in France, for coordinating the Young Boost Trial in France and the following for their active participation: M. van Hezewijk, Leiden (NL); M.J.C. van der Sangen Eindhoven (NL); M.C. Stenfert Kroese, Deventer (NL); J.J.Jobsen, Enschede (NL); J.M. Immink, Delft (NL); M.E. Mast, Den Haag (NL); F.M. Gescher, Den Haag (NL); N. Bijker, Amsterdam (NL); J.W.M. Mens, Rotterdam (NL); W.G.J.M. Smit, Leeuwarden (NL); D.H.F. Rietveld, Amsterdam (NL); I. Lecouillard, Rennes (Fr); C Breton-Callu, Bordeaux (Fr); H. Marsiglia, Villejuif (Fr); J. Thariat, Nice (Fr); A. Benyoucef, Rouen (Fr); A. Labib, Saint Cloud (Fr); M. Aumont, Saint Herblain (Fr); P. Bontemps, Besancon (Fr); C. Le Foll, Lagny (Fr); Y. Belkacemi, Créteil (Fr); O. Chapet, Lyon (Fr); V. Strnad, Erlangen (De)

5.5 Supplementary Material

(**a**) Right breast (**b**) Left breast

Figure 5.3: Coding of the tumour location

The dominant location of the lesion was represented by a number (figure 5.3). We lumped several regions together to create the different tumour locations: central under, lateral en medial upper:
Central under: 13, 14, 15 and 18 (right breast) and 23, 24, 25 and 28 (left breast).
Lateral: 11 and 19 (right breast) and 21 and 29 (left breast).
Medial-up: 12, 16 and 17 (right breast) and 22, 26 and 27 (left breast).

Table 5.6: Protocol violations

Major violations:		
	4	higher tumour stage than allowed
	2	residual microcalcifications on the post-operative mammography
	3	mastectomy
	2	different pathology
	1	withdrawn of patients' consent
	1	multifocal tumour
	1	no baseline photograph
	1	neoadjuvant chemotherapy
Total:	15	Major violations
Minor violations:		
	2	informed consent was received too late
	6	delay in start of radiation therapy after surgery
	1	51 years old
Total:	19	Minor violations
Unknown significance:		
	4	released by the investigator without giving a reason
	8	no specification of violation of inclusion criteria
Total:	12	Unspecified violations
Grand Total:	46	

Radiotherapy of the breast

The overall treatment time was kept constant in both randomization arms, i.e. 6.5–7 weeks. Fraction size was between 1.8 Gy per fraction and 2.3 Gy per fraction.

Radiation schedules: Standard radiation of the whole breast was 25×2 Gy, 5 fractions/week to the whole breast, followed by:

Low dose boost: 8×2 Gy = 16 Gy to the boost volume, 5 fractions/week, given after the whole breast irradiation.

High dose boost: 5×2 Gy to the boost volume, 1 fractions/week during the first 5 weeks, given with an interval of preferably 8 hours, but at least 6 hours with the radiation to the whole breast. Followed by 8×2 Gy to the boost volume, 5 fractions/week, given after the whole breast irradiation In case an integrated boost (SIB) was used, the Normalized Total Dose for tumour should be kept equal to 66 and 76 Gy for the low and high dose arm, respectively, using an α/β ratio of 10 Gy for the tumour. Low dose arm: 28 fractions of 1.81 Gy to the whole breast and an additional 0.49 Gy/fraction to the boost volume (total 2.3 Gy/fraction to the boost volume) in the same 28 fractions. High dose arm: 31 fractions of 1.66 Gy to the whole breast and an additional 0.72 Gy/fraction to the boost volume (total of 2.38 Gy/fraction to the boost volume) in the same 31 fractions

Low and high dose boost with interstitial radiation: The low dose boost was 15 Gy if given with LDR or PDR or 7.5 Gy if given with HDR. The high dose boost was 25 Gy if given with LDR or PDR or 10 Gy if given with HDR.

Target volume of the whole breast irradiation The Clinical Target Volume (CTV) includes the whole mammary gland. The mammary was carefully delineated (e.g. with a lead wire) at time of the simulation. For the Planning Target Volume (PTV) a margin of at least 0.5 cm is added to the CTV in dorsal and in craniocaudally direction. The skin (5 mm beneath the epidermis) was excluded from the PTV.

Boost target volume The boost CTV was in principle the rim of tissue 1.5 cm around the original tumour. The boost area was reconstructed using all available information, as preoperative physical examination, imaging and perioperative clips placed by the surgeon. The boost CTV was not to be extended into the skin (5 mm beneath the epidermis). In case of external beam irradiation an additional margin of 0.5 cm was added to create the PTV. The PTV was also not to be extended into the skin.

Treatment planning

Whole breast irradiation Patients were irradiated in supine position by two tangential fields. The dorsal borders of both fields were medially at least 0.5 cm outside the contour of the breast and laterally at least 1.5 cm outside the breast for thin patients. In patients with

a large amount of lateral subcutaneous fat, the distance of the dorsal border to the lateral breast contour should increase proportionally. In cranial-caudal direction the fields should have at least 1 cm margin between the edge of the field and the breast tissue. Wedge filters or compensators were advised for acceptable dose distribution. 60Co gamma-ray beams with a minimum SSD of 80 cm, or X-ray beams in the range of 4–18 MV were allowed.

Boost irradiation External beam irradiation was performed with two or more external photon beams or one electron beam. Irradiation of the contralateral breast was avoided as much as possible. Interstitial therapy: manual or remote control afterloading with Iridium-192 wires and stepping source afterloaders with Ir192 were allowed; Caesium-137 or Radium-226 needles were avoided. Experimental application with intracavitary balloons was not allowed.

Dose specifications $\leq 5\%$ of the CTV received $\geq 110\%$ of the prescribed dose (not for interstitial implants). $\leq 1\%$ of the CTV received $< 90\%$ of the prescribed dose (not for electron beam boosts). If no 3-dimensional dose distribution was available, the dose to the whole breast was specified at the ICRU reference point (point A), in agreement with ICRU report 50 and 62. This point was defined as the intersection of the beam axes. In case of an external photon boost, the boost dose was specified at point B (centre of the boost PTV). In case of an electron boost, the boost dose was specified at Dmax: the 85% isodose encompassed the boost PTV. In case of an interstitial boost, the dose was prescribed at the 85% peripheral isodose of the mean central dose (according to the ICRU report 58).

Organs at risk Dose to the normal tissues were kept as low as possible.

CHAPTER 6

Very Low Local Recurrence Rates After Breast-Conserving Therapy: Analysis of 8485 Patients Treated Over a 28-Year Period

This chapter was published as:
S. C. J. Bosma*, F. van der Leij*, E. van Werkhoven*, H. Bartelink, J. Wesseling, S. Linn, E. J. Rutgers, M. J. van de Vijver, and P. H. M. Elkhuizen.
Very low local recurrence rates after breast-conserving therapy: analysis of 8485 patients treated over a 28-year period.
Breast Cancer Research and Treatment, 156(2):391–400, 2016.
*) Equally contributing

Abstract

The purpose of this study was to study the impact of changes in clinical practice on outcome in patients treated with breast-conserving therapy (BCT) over a period of 28 years.

Patients with early invasive breast cancer, who were treated with BCT at the Netherlands Cancer Institute between 1980 and 2008, were studied. Clinical characteristics, treatment and outcome were compared between groups (1980–1987; 1988–1998; 1999–2008). The main endpoint analyzed was ipsilateral breast tumor recurrence (IBTR).

8485 patients with a median follow-up of 9 years (IQR 6–14 years) were analyzed. The cumulative 5- and 10-year IBTR incidences were, respectively, 2% and 5% for the whole cohort and 4% and 9% in patients $\leq$ 40 years. Young age was a significant risk factor for IBTR in multivariable analysis. IBTR-free interval was better for patients who received a RT boost (HR 0.65) or systemic therapy (HR 0.52). In later years, patients less often received a boost and more often underwent adjuvant systemic treatment. 761 patients (9.0%) underwent a re-excision; the tumor resection margins were tumor free for 85%. In later years (1999–2008), 89% of patients had a tumor-free margin. The margin status of invasive carcinoma did not influence IBTR, DM rate, or OS. Between 1980 and 2008, locoregional control after BCT remained stable with low IBTR rates, even in young patients.

These good results were achieved under the policy of accepting close or focally positive margins, indicating this is a safe approach. The results of this study may help in lowering the re-excision rates, which are high in many centers.

6.1 Introduction

Breast-conserving therapy (BCT) is currently the standard treatment for most patients with early-stage breast cancer. After publication of multiple prospective, randomized trials, which showed equivalent survival between BCT and mastectomy, the use of BCT has rapidly increased since the 1980s (Litière et al., 2012; Fisher et al., 2002; Veronesi et al., 2002; Van Dongen et al., 2000). In the past decades, major changes in the diagnosis and treatment of breast cancer took place: the introduction of population-based screening and the innovations in radiologic diagnostic methods, surgical techniques, pathologic analyses, and radiotherapy (RT) planning and treatment (Poortmans et al., 2012). Furthermore, the use of adjuvant systemic treatment has increased substantially and new drugs have been introduced, such as aromatase inhibitors, taxanes, and Trastuzumab (Harlan et al., 2006; Sukel et al., 2008; Perez et al., 2014; Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2012). Multiple studies have shown improved clinical outcome after BCT over time in terms of locoregional and distant recurrences and overall survival (Poortmans et al., 2012; Cabioglu et al., 2005; Van der Leest et al., 2007; Louwman et al., 2008).

The EBCTCG meta-analysis (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2011) demonstrated that whole breast irradiation (WBI) halves the breast-recurrence rate after wide local excision and also found an association between locoregional control and breast cancer mortality. Therefore, prevention of locoregional recurrence remains important.

The goal of the present study was to investigate time trends in BCT and the impact of these treatment changes on outcome of patients treated at The Netherlands Cancer Institute (NKI). For this purpose, we studied a large cohort of 10,509 early breast cancer patients, treated with BCT, at the NKI between 1980 and 2008. The specific objectives were to evaluate trends in locoregional and distant recurrences and overall survival and to identify factors with prognostic value.

6.2 Patients and methods

Study population

Patient data were obtained from the NKI's medical registry. 10,509 patients who were treated with surgery and/or RT and/or systemic treatment at the NKI as part of BCT between 1980 and 2008 were identified. Surgery was performed at the NKI (for approximately 70% of patients) or in surrounding hospitals. After 2007, follow-up data were unavailable for patients who received surgery elsewhere. Therefore, these patients were excluded ($n = 708$).

Other exclusion criteria were no RT, cM+, a previous/synchronous malignancy (< 90 days between the dates of diagnosis), non-invasive carcinoma, T3-T4 tumors, and neoadjuvant treatment (figure 6.1).

Stage was based on the pathological tumor-node-metastasis (TNM) classification. If pathological T stage could not be acquired, clinical T stage based on imaging was used ($n = 517$). Since 2002, information about the sentinel node (SN) was available to classify the lymph node stage in case an axillary lymph node dissection was not performed. We reclassified the

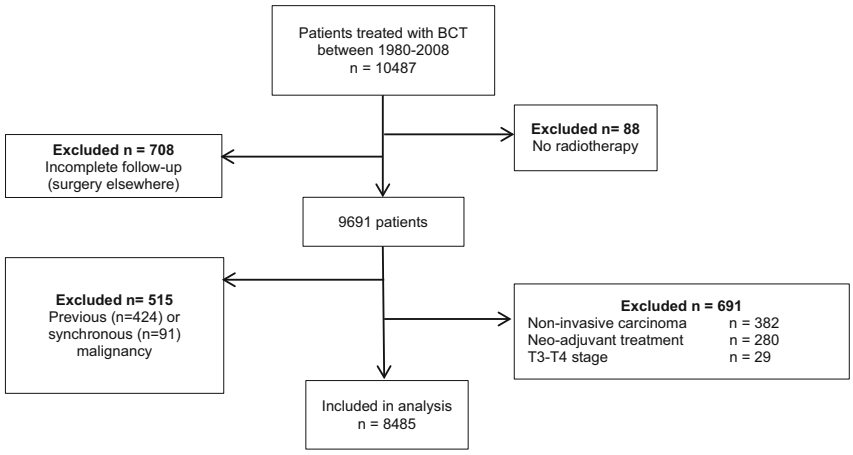


Figure 6.1: Flow diagram of the study population

axillary status as follows: SN negative as pN0, SN with isolated tumor cells as N0, SN with a micro metastasis (> 0.2 cm) as N1 mic, and SN with a metastasis as N1.

Pathology

When pathology information was missing from the medical registry this was retrieved from the PALGA system (Dutch Pathology Registry); estrogen and progesterone receptor status and human epidermal growth factor receptor-2 status were routinely reported since 2005. We added margin status of the invasive carcinoma and of the DCIS component, when present. We divided margin status into free, close (invasive tumor at 0.1–2.0 mm from the resection margin), focally positive (tumor in the inked resection margin over a distance ≤ 4 mm), and tumor positive. In earlier years, detailed margin status was not routinely assessed. According to the guidelines for BCT in the Netherlands close or focally positive margins are not an indication for re-excision as part of BCT. Since 2002, the Dutch National Guideline recommends to restrict re-excision to 'more than focally positive' resection margin (NABON, 2012).

Treatment

Treatment consisted of breast-conserving surgery and nodal staging with imaging, axillary dissection or SN procedure, followed by WBI with or without a boost. RT generally consisted of 50 Gray (Gy) in fractions of 2 Gy. An additional boost to the tumor bed could be given by external beam irradiation or by 192Ir Implantation (n = 1926), the latter being a frequent treatment modality in the earlier period covered by this study. We subdivided the boost dose into low (total dose 55–66 Gy) and high (total dose 67–80 Gy). In the first year of the study

period, a standard (high) boost dose of 26 Gy was given, when a boost was indicated. In later years, the boost dose was lowered to 16 Gy for patients with a microscopically complete excision and a boost of 26 Gy was only given in case of a microscopically incomplete excision or when an extensive DCIS component was present (Borger et al., 1994). During the study period, several clinical radiotherapy trials were running in which patients participated: the EORTC 10801 trial (1980–1986) (Litière et al., 2012), the boost-no boost EORTC 22881-10882 trial (1989–1996) (Poortmans et al., 2008), and the young boost trial (> 2004) for patients ≥ 50 years (`https://clinicaltrials.gov/ct2/show/NCT00212121`).

Adjuvant systemic therapy was given according to the Dutch National Guidelines or trials. In the course of the study period, these guidelines have changed, and we subdivided the study population accordingly (1980–1987, 1988–1998, and 1999–2008). Before 1982, systemic therapy was not advised for early-stage breast cancer. Thereafter, systemic therapy was introduced in an adjuvant setting for patients with regional lymph nodal metastases. The hospital guidelines advised chemotherapy for premenopausal patients and no adjuvant therapy for postmenopausal patients. After 1990, node-positive patients < 50 years received adjuvant chemotherapy and patients > 50 years received adjuvant hormonal therapy (usually tamoxifen for 5 years). After publication of a new guideline of the Dutch National Breast Cancer Platform and the Dutch Society for Medical Oncology in 1998, adjuvant systemic therapy was also advised for high risk (based on primary tumor characteristics) node-negative patients (Bontenbal et al., 2000). We classified adjuvant systemic treatment into chemotherapy, hormonal therapy (Tamoxifen and aromatase inhibitors), and immunotherapy (Trastuzumab). The standard chemotherapy regimen in the beginning of the study period consisted of cyclophosphamide, methotrexate, and fluorouracil (CMF), thereafter an anthracycline-containing regimen. Trastuzumab with a taxane has been added to adjuvant anthracycline-based chemotherapy since 2005.

Statistical Analysis

The endpoints of this study were time to ipsilateral breast tumor recurrence (IBTR), distant metastases (DM), and overall survival (OS) since date of diagnosis. IBTR was defined as any recurrence of invasive breast cancer in the ipsilateral breast or chest wall or regional lymph nodes. Only IBTRs that were a first and isolated event (no other event within 3 months) were taken into account. Patients with contralateral breast cancer and non-breast second primary cancer, or patients who died without disease recurrence were censored. Survival curves were estimated using the Kaplan-Meier method. Patients with follow-up longer than 20 years were censored at 20 years for the analysis of IBTR-free interval, since follow-up beyond 20 years was unreliable with only limited number of patient data available. Multivariable analyses were performed using Cox proportional hazards models. Missing values were coded as a separate category in order to keep all patients in the model. The prognostic impact of IBTR and DM on OS was studied using time-dependent variables. Differences in baseline characteristics between the different groups, based on treatment period, were tested (table 6.1) and p values of 0.05 or less were considered significant.

6.3 Results

Patient, tumor, and treatment characteristics

8485 consecutive patients with early invasive breast cancer who were treated with BCT between 1980 and 2008 were analyzed (table 6.1). The median follow-up was 9 years (IQR 6–14 years).

Table 6.1: Patient characteristics per treatment period

	1980–1987 1068 (13%)	1988–1998 3612 (42%)	1999–2008 3805(45%)	Total 8485
Age at diagnosis ($p < 0.0001$)				
Median (range)	49 (22–90)	53 (23–89)	56 (20–90)	54 (20–90)
Age category				
[20,41)	228 (21%)	435 (12%)	300 (8%)	963 (11%)
[41,51)	363 (34%)	1070 (30%)	881 (23%)	2314 (27%)
[51,65)	339 (32%)	1308 (36%)	1622 (43%)	3269 (39%)
[65,75)	114 (11%)	618 (17%)	745 (20%)	1477 (17%)
[75,90]	24 (2%)	181 (5%)	257(7%)	462 (5%)
Tumor type				
IDC	684 (64%)	2691 (75%)	3028 (80%)	6403 (75%)
ILC	86 (8%)	320 (9%)	383 (10%)	789 (9%)
Other	285 (27%)	397 (11%)	249 (7%)	931 (11%)
mixed IDC + ILC	13 (1%)	204 (6%)	145 (4%)	362 (4%)
Localisation ($p < 0.001$)				
Central/medial	499 (47%)	1476 (41%)	1492 (39%)	3467 (41%)
Lateral/Axillary tail	566 (53%)	2131 (59%)	2305 (61%)	5002 (59%)
NA	3	5	8	16
T stage (p) ($p < 0.001$)				
1	719 (80%)	2561 (73%)	2888 (76%)	6168 (75%)
2	183 (20%)	953 (27%)	903 (24%)	2039 (25%)
NA	166	98	14	278
N stage				$p = 0.025$
Negative	711 (70%)	2245 (67%)	2008 (62%)	4964 (65%)
Positive	301 (30%)	1014 (30%)	942 (29%)	2257 (30%)
Micro	4 (0%)	81 (2%)	302 (9%)	387 (5%)
NA	52	272	553	877
Histological grade ($p < 0.001$)				
1	73 (15%)	643 (24%)	903 (24%)	1619 (23%)
2	166 (34%)	1201 (45%)	1847 (50%)	3214 (47%)
3	251 (51%)	850 (31%)	960 (26%)	2061 (30%)
NA	578	918	95	1591
ER status				
Negative	49 (29%)	484 (23%)	614 (17%)	1147 (20%)
Positive	120 (71%)	1642 (77%)	2953 (83%)	4715 (80%)
NA	899	1486	238	2623

(Continued on next page)

Table 6.1 – continued from previous page

	1980–1987 1068 (13%)	1988–1998 3612 (42%)	1999–2008 3805(45%)	Total 8485
PR status				
Negative	29 (29%)	430 (32%)	946 (35%)	1405 (34%)
Positive	72 (71%)	906 (68%)	1720 (65%)	2698 (66%)
NA	967	2276	1139	4382
HER2-neu status				
Negative	0	129 (90%)	1741 (86%)	1870 (86%)
Positive	0	14 (10%)	280 (14%)	294 (14%)
NA	1068	3469	1784	6321
Margins invasive ca				
Free	234 (69%)	2556 (81%)	3266 (89%)	6056 (85%)
Not free*	105 (31%)	583 (19%)	401 (11%)	1089 (15%)
NA	729	473	138	1340
Margin DCIS				
Free	11 (31%)	263 (48%)	464 (60%)	738 (55%)
Not free*	24 (68%)	283 (52%)	304 (40%)	611 (45%)
NA	1033	3066	3037	7136
Any Systemic therapy				
No	866 (81%)	2146 (59%)	1848 (49%)	4860 (57%)
Yes	202 (19%)	1466 (41%)	1957 (51%)	3625 (43%)
Hormonal therapy	34 (3%)	1023 (28%)	1510 (40%)	2567 (30%)
Chemotherapy	171 (16%)	551 (15%)	1136 (30%)	1858 (22%)
Boost dose				
No boost	17 (2%)	208 (6%)	1462 (42%)	1687 (21%)
Low boost (16 Gy)	304 (29%)	2731 (76%)	1759 (50%)	4794 (59%)
High boost (26 Gy)	729 (69%)	638 (18%)	263 (8%)	1630 (20%)
NA	18	35	321	374

*) including Focally positive / close

761 of the total 8485 patients (8.9%) underwent a re-excision. Re-excisions were performed in 10% of the patients from 1980 to 1999 and in 7.6% from 1999 to 2008. Margin status before re-excision was known for 6810 patients: in 5458 (80%) patients the tumor was excised with a free, in 197 (3%) with a focally irradical resection margin, in 621 (9%) with a close resection margin, and 534 (8%) with a more than focally tumor-positive resection margin. Final (after potential re-excision) margins for invasive carcinoma are displayed in table 6.1. In 3763 patients, a DCIS component was present. In 1349 patients, margin status of the DCIS component was reported (table 6.1).

Over time, changes in initial presentation included an increase in median age, increase in lateral tumors, and a decrease in the proportion of patients with a high-grade tumor. The percentage of patients presenting with micrometastases increased over time, as a result of the introduction of the SN procedure. In later periods, more T2 tumors were treated (table 6.1).

6424 of 8485 patients (79%) received boost irradiation additional to WBI: of which

4794 (75%) a low (16 Gy) and 1630 (25%) a high (26 Gy) boost. Young patients more often received a boost and a higher boost compared to older patients: 4% of patients ≤ 40 years received no boost versus 27% of patients aged 51–64 years; a high boost was given to 33% of patients ≤ 40 years versus 17% of patients aged 51–64 years. The proportion of patients receiving a boost declined over time, especially the proportion receiving a high boost (table 6.1). The decline was mostly seen in older patients (≥ 65 years). In this patient group, 3% and 5% received no boost between 1980 and 1987 and between 1988 and 1998, respectively, and this increased to 69% between 1999 and 2008.

3625 of 8485 patients (43%) received adjuvant systemic therapy. Treatment with adjuvant systemic therapy, both chemotherapy and hormonal therapy, increased over time (table 6.1). In later years, adjuvant systemic therapy was more frequently given to node-negative patients: 1.5% of node-negative patients received adjuvant systemic therapy between 1980 and 1987, 13% between 1988 and 1998, and 33% between 1999 and 2008. The percentage of young patients (≤ 40 years) treated with adjuvant systemic therapy increased from 29% between 1980 and 1987 to 35% between 1988 and 1998 to 81% between 1999 and 2008.

Ipsilateral breast tumor recurrence (IBTR)

The 5- and 10-year IBTR incidences were, respectively, 2% (95% CI 2–3%) and 5% (95% CI 5–6%) for the whole cohort, and significantly higher, respectively, for younger patients (≤ 40 years): 4% (95% CI 3–6%) and 9% (95% CI 7–12%) (figure 6.2). IBTR rates were similar between patients who had surgery at the NKI-AVL or elsewhere ($p = 0.2$).

The IBTR-free interval improved slightly, but significantly over time ($p = 0.045$). There was no interaction between age and time period of diagnosis ($p = 0.46$). Univariable analysis for IBTR was performed with the variables displayed in table 6.1. Significant risk factors for IBTR were early treatment period, young age, positive N stage, high histological grade, and focally incomplete/close/involved margins with DCIS; factors associated with a lower risk of IBTR were adjuvant systemic therapy and a RT boost. The type of RT boost did not significantly influence the IBTR risk. All factors were taken into account in multivariable analysis (table 6.2). Re-excision did not significantly affect the IBTR rate ($p = 0.66$) in patients with focally incomplete, close, or tumor-positive margins for invasive carcinoma ($n = 1089$).

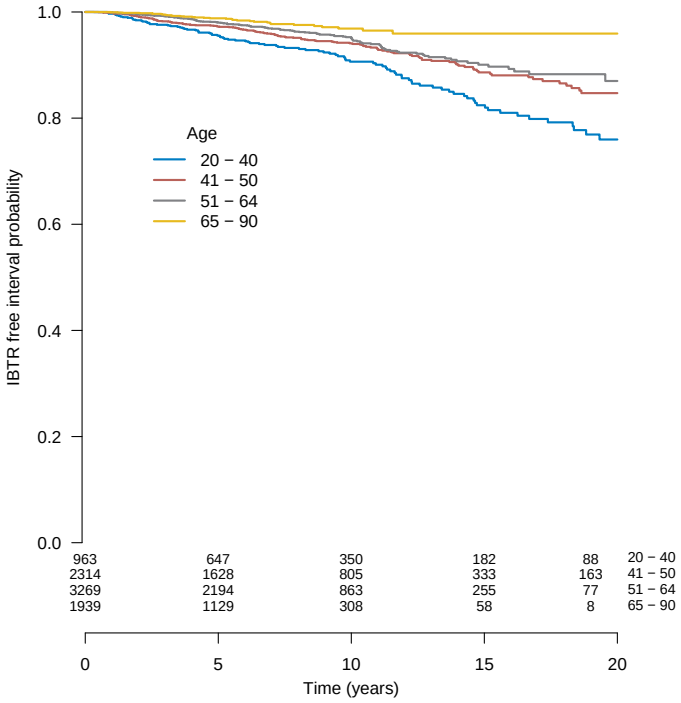


Figure 6.2: IBTR-free interval by age group

Table 6.2: Multivariable analysis for risk factors for IBTR

Variable	HR (95%CI)	p-value
Diagnosis time		
1980 - 1987	1	
1988 - 1998	1.18 (0.86–1.61)	0.31
1999 - 2008	0.89 (0.54–1.47)	0.67
Age at diagnosis		
[20,41)	1	
[41,51)	0.62 (0.48–0.82)	0.001
[51,65)	0.51 (0.38–0.68)	< 0.0001
[65,90]	0.28 (0.18–0.43)	< 0.0001
Localisation		
Central / medial	1	
Lateral / Axillary tail	0.93 (0.76–1.14)	0.49

(Continued on next page)

Table 6.2 – continued from previous page

Variable	HR (95%CI)	*p*-value
T stage		
1	1	
2	1.26 (0.98–1.61)	0.07
Missing	0.90 (0.50–1.45)	0.74
N stage		
Negative	1	
Micro	0.50 (0.20–1.24)	0.14
Positive	1.37 (0.98–1.92)	0.06
Missing	1.07 (0.73–1.59)	0.72
ER status		
Negative	1	
Positive	0.97 (0.67–1.41)	0.89
Missing	0.76 (0.50–1.17)	0.21
PR status		
Negative	1	
Positive	1.01 (0.69–1.48)	0.96
Missing	0.99 (0.65–1.48)	0.74
HER2 status		
Negative	1	
Positive	1.08 (0.48–2.41)	0.86
Missing	0.88 (0.58–1.34)	0.56
Differentiation grade		
Grade 1	1	
Grade 2	1.71 (1.22–2.42)	0.002
Grade 3 / Anaplastic	1.47 (0.99–2.19)	0.06
Missing	1.61 (1.11–2.32)	0.01
Margin invasive ca		
Free	1	
Not free / close / focally irradical	0.97 (0.73–1.29)	0.83
missing	1.10 (0.83–1.45)	0.51
Margin DCIS		
free	1	
Not free	2.28 (1.27–4.07)	0.01
No DCIS	1.10 (0.64–1.88)	0.73
NA	1.47 (0.88–2.44)	0.14
Chemotheraly		
No	1	
Yes	0.57 (0.40–0.83)	0.003

(Continued on next page)

Table 6.2 – continued from previous page

Variable	HR (95%CI)	p-value
Hormonal therapy		
No	1	
Yes	0.48 (0.34–0.67)	< 0.0001
RT boost		
No boost	1	
Boost	0.62 (0.45–0.86)	0.004
Missing	0.68 (0.32–1.44)	0.31

Patients with an IBTR had a significantly higher risk of developing DM (HR 5.21; 95% CI 4.3–6.3; $p < 0.0001$) and significantly worse OS (HR 1.86; 95%CI 1.6–2.2; $p < 0.0001$) compared to patients without an IBTR. The median OS after the diagnosis of an IBTR was 11.4 years (95% CI 9.1–15.2 years), the 5- and 10-year OS rates were 74% (95%CI 69–79%) and 53% (95%CI 69–79%) respectively. OS was worse for patients with early IBTR (< 2 years) compared to patients with late IBTR (> 2 years) ($p < 0.0001$, figure 6.3) and for patients > 50 years compared to patients ≤ 50 years ($p < 0.0001$). Over time, no improvement of OS after IBTR was observed ($p = 0.37$).

Distant metastases (DM)

1306 patients developed distant metastases. The 5-and 10-year incidences of DM were 11% (95% CI 10–12%) and 18% (95%CI 17–19%), respectively.

Multivariate analysis for DM was performed with the same factors as multivariate analysis for LR (see addendum for multivariable analysis of risk factors for DM and OS). DM-free interval was better in patients treated more recently (HR 0.57; 95%CI 0.4–0.8; $p = 0.001$ for patients treated between 1999 and 2008 compared to patients treated between 1980 and 1987). Younger patients (≤ 40 years) more often developed DM compared to older patients (HR 0.71; 95%CI 0.6–0.9; $p < 0.001$ for patients of 51–64 years old).

Other risk factors for developing DM were: higher tumor stage (T2 vs T1 stage HR 1.74; 95% CI 1.5–2.0, $p < 0.0001$), positive N status (positive vs negative N status HR 2.10; 95%CI 1.8–2.5; $p < 0.001$), and poor histologic grade (grade 3 vs grade 1 HR 3.98; 95%CI 3.0–5.3; $p < 0.0001$). Patients with a lateral tumor had a lower risk of DM compared to patients with a medial tumor (HR 0.75; 95%CI 0.7–0.9; $p < 0.0001$). The use of chemotherapy reduced the risk of DM (HR 0.76; 95%CI 0.6–0.9; $p < 0.01$). The median OS time after the detection of DM was 2.0 years (95%CI 1.8–2.1 years) and the 5- and 10-year overall survival rates were 18% (95%CI 16–20%) and 5.4% (95%CI 4.1–7.0%), respectively.

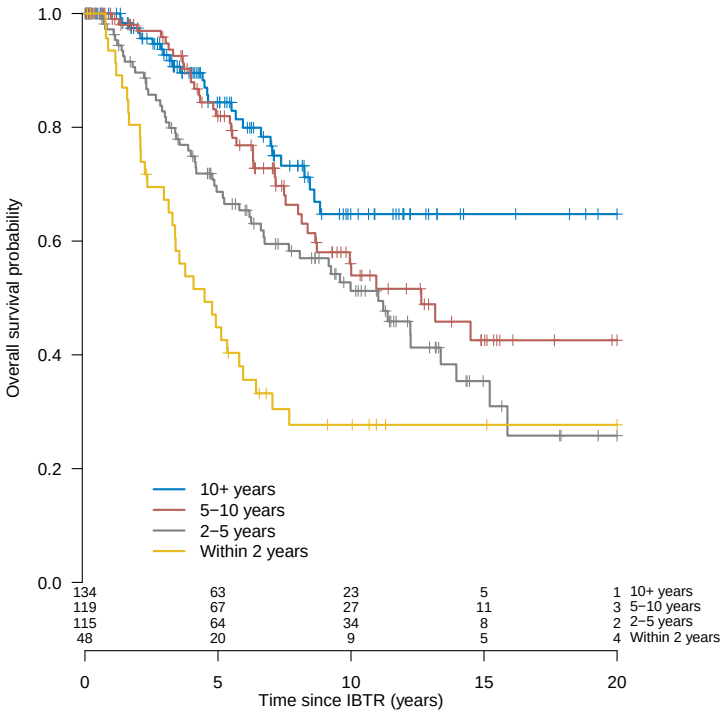


Figure 6.3: Overall survival after IBTR, grouped by time from diagnosis to IBTR

Overall Survival

2174 of the patients had died, and death was breast cancer related in 1078 patients. OS was 91% (95%CI 90–92%) at 5 years and 77% (95% CI 76–78%) at 10 years. On multivariable analysis (table 6.3), significant risk factors for worse OS were higher age (only for patients 65–90 HR 2.71; 95%CI 2.2–3.2; $p < 0.0001$ compared to patients ≤ 40 years old), higher tumor stage (T2 vs T1 stage HR 1.54; 95%CI 1.4–1.7; $p < 0.0001$), positive N status (positive vs negative N status HR 1.72; 95%CI 1.5–2.0; $p < 0.0001$), and poor differentiation grade (grade 3 vs grade 1 HR 2.10; 95%CI 1.8–2.5; $p < 0.0001$).

Patients with a lateral tumor had better OS compared to patients with a medial tumor (HR 0.88; 95%CI 0.8–0.98; $p = 0.02$). Chemotherapy was associated with better OS (HR 0.84; 95% CI 0.7–1.0; $p = 0.05$ for patients treated with chemotherapy compared to no chemotherapy). Systemic therapy resulted in a greater improvement of OS in patients with positive lymph nodes compared to patients with negative lymph nodes (interaction p-value 0.006). Patients treated between 1999 and 2008

had better OS compared to patients treated in earlier time periods (table 6.3).

Table 6.3: Multivariable analysis of distant metastasis free (DM) and overall survival (OS) time

	DM		OS	
	HR (95%CI)	p-value	HR (95%CI)	p-value
Diagnosis time				
1980—1987	1		1	
1988–1998	0.98 (0.8–1.2)	0.88	1.19 (1.0–1.4)	0.07
1999–2008	0.57 (0.4–0.8)	0.001	0.96 (0.7–1.2)	0.77
Age				
20–41	1		1	
41–51	0.77 (0.6–0.9)	0.01	0.94 (0.7–1.1)	0.51
51–65	0.71 (0.6–0.9)	0.001	1.21 (1.0–1.5)	0.06
65–90	0.63 (0.5–0.8)	0.0005	2.71 (2.2–3.2)	< 0.0001
Localisation				
Medial	1		1	
Lateral	0.75 (0.7–0.9)	< 0.0001	0.88 (0.8–0.98)	0.02
T-stage				
1	1		1	
2	1.74 (1.5–2.0)	< 0.0001	1.54 (1.4–1.7)	< 0.0001
N-stage				
Negative	1		1	
Micro	0.97 (0.6–1.5)	0.9	0.98 (0.7–1.4)	0.92
Positive	2.10 (1.8–2.5)	< 0.0001	1.72 (1.5–2.0)	< 0.0001
ER-status				
Negative	1		1	
Positive	1.13 (0.9–1.4)	0.33	0.91 (0.8–1.1)	0.36
PR-status				
Negative	1		1	
Positive	0.73 (0.8–0.9)	0.01	0.91 (0.7–1.1)	0.34
HER2-status				
Negative	1		1	
Positive	0.68 (0.4–1.1)	0.13	0.89 (0.6–1.3)	0.6
Differentiation grade				
1	1		1	
2	2.53 (1.9–3.3)	< 0.0001	1.39 (1.2–1.6)	0.0001
3	3.98 (3.0–5.3)	< 0.0001	2.10 (1.8–2.5)	< 0.0001
Margin invasive ca				
Free	1		1	
Irradical/Not Free	1.01 (0.8–1.3)	0.92	1.00 (0.8–1.2)	1
Margin DCIS				
Free	1		1	
Irradical/Not free	1.16 (0.8–1.7)	0.42	1.18 (0.9–1.6)	0.23
no DCIS	0.98 (0.7–1.3)	0.88	0.91 (0.8–1.3)	0.91
Chemotherapy				
No	1		1	
Yes	0.76 (0.6–0.9)	0.01	0.84 (0.7–1)	0.05

(Continued on next page)

Table 6.3 – continued from previous page

	DM		OS	
	HR (95%CI)	p-value	HR (95%CI)	p-value
Hormonal therapy				
No	1		1	
Yes	0.89 (0.7–1.1)	0.21	0.89 (0.8–1.1)	0.12
RT boost				
No	1		1	
Yes	0.98 (0.8–1.2)	0.89	1.12 (0.9–1.3)	0.22

*) including Focally positive / close

6.4 Discussion

Low 5- and 10-year IBTR incidences of, respectively, 2% and 5% were observed in this cohort of 8485 patients treated with BCT between 1980 and 2008 at the Netherlands Cancer Institute. Even in young patients (≤ 40 years), who are known to be at relatively high risk of IBTR in many studies (Bartelink et al., 2015; Elkhuizen et al., 1998; Bollet et al., 2007; Vrieling et al., 2003), a low IBTR rate was found of 4% and 9% after, respectively, 5 and 10 years. Already in the earlier years of the time covered in this study, low IBTR incidences were achieved with a high radiation dose and limited use of systemic therapy.

For comparison, in the EORTC 10801 trial (1980–1986), the 10-year IBTR rate was 20% (Van Dongen et al., 2000). In successive EORTC trials concerning BCT, the 10-year IBTR rates declined to 10% between 1989 and 1996 (boost arm of the boost-no boost trial) (Van der Leest et al., 2007) to $< 2\%$ at 8 years in patients treated since 2004 (Young Boost) (Bartelink et al., 2012). In our cohort, a small improvement of locoregional control over time was observed. However, after correction for unfavorable tumor characteristics in earlier time periods, this time trend disappeared. As the role of adjuvant systemic therapy was limited in the earlier time periods, the good local control in general clinical practice should be attributed to other factors, such as optimal imaging, surgery, and histopathological assessment. An important factor reducing IBTR in our cohort is the frequent use of a boost of radiotherapy.

The risk of IBTR was reduced by the use of a RT boost by 35% ($p = 0.01$), which is in accordance with the EORTC boost-no boost trial (Bartelink et al., 2015). The use and the dose of the RT boost declined over time in the present study. Nevertheless, the IBTR rate remained stable, which can most likely be explained by the increased use of systemic therapy (from 19% between 1980 and 1987 to 54% between 1999 and 2008). This was associated with a reduced risk of local and distant recurrences and death. Multiple studies have shown that improved locoregional control was associated with increased adjuvant systemic therapy (Cabioglu et al., 2005; Van der Leest et al., 2007; Van Laar et al., 2013; Bouganim et al., 2013), although it may also

be due to improved patient selection and local treatment (Ernst et al., 2004).

During the past decades, there have been changes in different aspects of the multidisciplinary treatment of BCT. The population-based mammographic screening was introduced in the Netherlands in 1990 for patients aged 50–70 years and became available for patients from 70 to 75 years in 1997, which led to a changing case mix due to a shift toward tumors with favorable prognostic clinico-pathological factors and better prognosis (Mook et al., 2011).

In the field of pathological examination, margin assessment has greatly improved. In earlier years, this was not routinely performed. Remarkably, for patients treated during this time period similar IBTR rates were achieved compared to later years. In our cohort, 9.0% of patients underwent a re-excision. In the latest period studied in this study, which is most representative for current daily clinical practice, 7.6% of all patients underwent a re-excision and 89% had a tumor-free margin. Similar results were shown in a study of 7345 Dutch patients treated in various hospitals in which 9.1% of patients with invasive tumor had 'more than focally positive' tumor margins (Van der Heiden-van der Loo et al., 2012). In a subgroup analysis of the EORTC 'boost-no boost' trial close or tumor-positive margin status did not significantly influence local control (Jones et al., 2009). However, in some studies, an association between margin status and local control was observed. While the optimal margin width is unclear, 'no ink on tumor' is accepted as an adequate margin (Houssami & Morrow, 2014). But varying guidelines for re-excision are used.

Since 2002, the Dutch National Guideline recommends to restrict re-excision to 'more than focally positive' resection margin or multiple focal areas. This is in contrast to, e.g., the United States of America where high re-excision rates of 19 and 23% in large multi-institutional studies have been reported (Morrow et al., 2009; McCahill et al., 2012). In addition, positive margin rates are higher. For example, in a recent trial concerning cavity shaving after breast-conserving surgery, a high positive margin rate of 34% was observed (Chagpar et al., 2015). Our study shows that with less stringent re-excision guidelines, good local control can also be achieved. This was also observed in a study of 40,892 patients treated with BCT between 2003 and 2006 in the Netherlands, which showed an overall a 5-year IBTR rate of 2.85% (95%CI 2.68–3.03) (Van der Heiden-van der Loo et al., 2015). Potentially worse cosmetic outcomes, additional health care costs and stress for patients can be avoided by reducing re-excisions (Moran et al., 2014).

Patients who developed an IBTR had significantly worse DM-free interval (HR 5.21) and OS (HR 1.86) compared to patients who did not. Time to recurrence was a significant prognostic factor for OS, as also shown by others (Wapnir et al., 2006; Anderson et al., 2009). Other significant risk factors for developing DM were young age, larger tumors, central/medial localisation, positive N status, and high histological grade. Except for young age, the same factors were associated with worse OS.

Patients with medial tumors had worse DM-free and OS compared to patients with a lateral tumor, but no inferior local control. Others previously demonstrated the negative influence of inner quadrant localization on OS (Gaffney et al., 2003). Estrogen receptor status was not a significant factor in our study, while some studies showed that a positive status was associated with better overall survival (Bentzon et al., 2008).

In many studies, young age ($\leq$ 40 years) is an independent risk factor for locoregional and distant recurrences after BCT. High 10-year IBTR rates ranging from 14 to 38% have been reported (Van der Leest et al., 2007; Bollet et al., 2007; Van Laar et al., 2013; Bartelink et al., 2007). Therefore, some doubt exists concerning the safety of BCT in young patients. Although younger age is an independent risk factor for IBTR in our study, the absolute IBTR risk was quite low (cumulative incidences of IBTR of 4% at 5 years and 9% at 10 years). Comparison of local control between studies is difficult because patient characteristics differ, and patients were treated in different time periods with varying treatment guidelines.

In older patients ($\geq$ 65 years), the 5- and 10-year IBTR rates were, respectively, even lower: 1 and 3%. For patients with low IBTR risk, several studies are undertaken to determine whether RT may be safely omitted with or without systemic therapy (Blamey et al., 2013; Hughes et al., 2013). In addition, as an alternative to whole breast irradiation partial breast irradiation is currently studied, since most of the local recurrences occur at or near the tumor bed (Sanders et al., 2007). Further, in decision making for the use of RT the biological background of breast cancer may play a role (Drukker et al., 2014).

Limitations of the current study are a consequence of its retrospective nature. The collection of follow-up data may not be complete. In earlier time periods, histopathological data were missing. We completed this information as much as possible and succeeded for the last time periods, which are most representative for current daily practice.

In summary, our study shows excellent locoregional control during the past decades even in younger women. This supports the use of BCT in all patient groups, especially when adjuvant systemic therapy is part of the treatment plan. Importantly, it shows that with less stringent re-excision guidelines and a subsequent low number of re-excisions, good local control can be achieved.

Acknowledgements

We thank the Medical Registry staff of the NKI and Stella Mook for providing baseline data.

CHAPTER 7

A Propensity Score-Adjusted Comparison of Lenalidomide + R-CHOP Versus R-CHOP for *MYC*-Rearranged DLBCL Patients

This chapter was submitted as:
Vera de Jonge*, Erik van Werkhoven*, Avinash G. Dinmohamed, Marcel Nijland, Aeilko H. Zwinderman, Patrick M. Bossuyt, Martine S. Veldhuis, Emma Rutten, Rogier Mous, Joost S.P. Vermaat, Yorick Sandberg, Eva de Jongh, Yavuz Bilgin, Rinske Boersma, Harry Koene, Marie José Kersten, Daphne de Jong, Martine E.D. Chamuleau
A Propensity Score-Adjusted Comparison of Lenalidomide + R-CHOP Versus R-CHOP for *MYC*-Rearranged DLBCL Patients.
*) Equally contributing

Abstract

Patients with *MYC* rearranged (*MYC*-R) diffuse large B-cell lymphoma (DLBCL) generally have a poor prognosis following standard treatment with R-CHOP. Previously, we demonstrated in a single-arm phase II trial (HOVON-130) that addition of lenalidomide to R-CHOP (R2CHOP) is well-tolerated and yields similar complete metabolic remission (CMR) rates as more intensive chemotherapy regimens in literature. In parallel with this single-arm interventional trial, a prospective observational screening cohort (HOVON-900) was open in which we identified all newly diagnosed *MYC*-R DLBCL patients in the Netherlands. Patients from the observational cohort who met the inclusion criteria for the interventional trial but were not included, were treated with standard R-CHOP. This cohort served as control group in the present propensity-score adjusted comparison. Clinical information at presentation and treatment and outcome data were retrieved from the Netherlands Cancer Registry.

Patients from the interventional trial treated with R2CHOP (n=77) were younger than patients in the control cohort (n=56) treated with R-CHOP (median age 63 versus 70 years, p=0.018) and they were more likely to have a lower WHO performance score (p=0.013). To adjust for these and other differences at baseline, we used three statistical methods (multivariable analysis, 1:1 matching and weighting using the propensity score) to reduce treatment-selection bias. These consistently showed improved outcome after R2CHOP with HRs of 0.53, 0.53, and 0.59, respectively, for OS, and 0.59, 0.53, and 0.60 for PFS. Thus, this propensity score-adjusted comparison study supports the addition of lenalidomide to R-CHOP as a valuable treatment option for *MYC*-R DLBCL patients.

Key Points:

- Lenalidomide plus R-CHOP (R2CHOP) improves survival of *MYC*-R DLBCL patients over R-CHOP in a propensity score-adjusted comparison.

7.1 Introduction

First-line immunochemotherapy with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) cures the majority of diffuse large B-cell lymphoma (DLBCL) patients (Coiffier et al., 2002; Habermann et al., 2006). The most commonly used prognostic score is the International Prognostic Index (IPI), which consists of age (> 60 years), Ann-Arbor stage (III/IV), WHO performance score (≥ 2), lactate dehydrogenase (LDH) serum level (elevated), and number of extra-nodal localizations (> 1) (Int. NHL Prognostic Factors Project, 1993; Sehn et al., 2007). Other well-known prognostic disease characteristics are sex (Yıldırım et al., 2015) and the presence of a *MYC* rearrangement (*MYC*-R), which is detected in 10–15% of all newly diagnosed DLBCL cases (Rosenwald et al., 2019). Compared with a 5-year overall survival (OS) of 72% and 5-year progression-free survival (PFS) of 66% in patients without a *MYC*-rearrangement, *MYC*-R patients have a 5-year OS and PFS of 33% and 31%, respectively (Savage et al., 2009). In 70% of *MYC*-R patients a *MYC* rearrangement is detected with a concomitant *BCL2* or *BCL6* rearrangements (double hit [DH]), or with both *BCL2* and *BCL6* rearrangements (triple hit [TH]) (Aukema et al., 2014). The remaining 30% of the patients only have a *MYC* rearrangement only (single hit [SH]) (Rosenwald et al., 2019). The inferior prognosis of a *MYC* rearrangement is largely attributed to patients with a DH/TH lymphoma (Rosenwald et al., 2019) and, therefore, these subsets have been defined as a separate entity since 2016 (Swerdlow et al., 2016).

Intensified immunochemotherapy regimens have been investigated to improve first-line treatment for *MYC*-R patients. Such regimens, e.g. hyper-CVAD and R-CODOX-M/R-IVAC, seemed to improve PFS in phase II studies, but no OS improvement was demonstrated (Oki et al., 2014). In another prospective study, dose-adjusted EPOCH-R (DA-EPOCH-R) showed promising complete metabolic remission (CMR) rates of 74% at end of treatment and resulted in a 4-year event-free survival (EFS) of 71% and OS of 77% for all *MYC*-R patients (Dunleavy et al., 2018). DH/TH patients had an even better EFS of 73% and OS of 82%. Based on this study, many groups worldwide consider DA-EPOCH-R as the preferred first-line regimen for *MYC*-R patients, especially for DH/TH patients.

Other strategies to improve outcome for *MYC*-R DLBCL patients have focused on addition of novel drugs to the R-CHOP backbone. For example, in the CAVALLI phase II study, the selective *BCL2* inhibitor venetoclax was added to R-CHOP showing promising results, especially in DH lymphomas with high levels of *BCL2* protein expression (Morschhauser et al., 2021). Adding venetoclax to DA-EPOCH-R , however, turned out to be too toxic, resulting in early discontinuation of the subsequent phase III randomized study in DH lymphomas (Abramson et al., 2021).

The rationale for adding lenalidomide to the R-CHOP backbone for *MYC*-R DLBCL is the *MYC* downregulating effect of lenalidomide via cereblon targeting (Lopez-

Girona et al., 2011; Gopalakrishnan et al., 2016). In a single-arm phase II trial for newly diagnosed *MYC*-R patients, we have shown that addition of lenalidomide to R-CHOP is well-tolerated and resulted in a complete metabolic remission (CMR) in 67% of patients at end of treatment and a 2-year OS and EFS of 73% and 63%, respectively (Chamuleau et al., 2020). Here, we have selected a cohort of *MYC*-R patients from a simultaneously open, prospective population-based registration cohort of R-CHOP-treated DLBCL patients (HOVON-900 cohort) as controls to compare with the long-term follow-up data of the R2CHOP interventional group (HOVON-130 trial). In this comparison, we use three statistical models (multivariable analysis, 1:1 matching of the groups on IPI score, and propensity score weighting) to assess the added value of lenalidomide to R-CHOP in terms of OS and PFS.

7.2 Methods

Patient selection

In the HOVON-130 trial, patients aged $\geq$ 18 years with *MYC*-R DLBCL were included and treated with R-CHOP21 plus lenalidomide 15 mg day 1–15 for 6 cycles (Chamuleau et al., 2020). Additional inclusion criteria were Ann Arbor stage II-IV, a WHO performance status of 0–3, $\geq$ one lesion of $\geq$ 1.5 cm on a contrast-enhanced CT scan and $\geq$ one FDG-positive lesion on PET-CT scan. Patients diagnosed with any other subtype of aggressive B-cell lymphoma, a history of follicular lymphoma, proven CNS localization, or HIV infection were excluded.

Concurrent with the HOVON-130 trial (2015-2019), the HOVON-900 observational protocol was open for registration of any newly diagnosed *MYC*-R DLBCL patients in the Netherlands (Chamuleau et al., 2017). *MYC*, *BCL2* and *BCL6* fluorescent in situ hybridization (FISH) diagnostics were advocated as part of routine procedures. The HOVON Pathology Facility performed the pathology reviews.

For the present study, we selected all HOVON-900 *MYC*-R DLBCL patients treated with R-CHOP and who met all inclusion criteria of the HOVON-130 trial. Clinical data at presentation as well as routinely collected outcome data were retrieved from the Netherlands Cancer Registry (NCR). The HOVON-130 trial (Eudra-CT 2014-002654-39) and the HOVON-900 cohort were approved by the medical ethics committee (METC) of the Amsterdam UMC (METC VUMC 2015.082). The NCR Privacy Review Board additionally approved the use of anonymous data for this study.

Statistical methods

The primary endpoint overall survival time (OS) and the secondary endpoint progression-free survival time (PFS) were calculated from date of diagnosis to

death (OS) and to relapse or death (PFS). Patients alive were censored for OS, and patients alive without progression or relapse were censored for PFS at last follow-up. The Kaplan-Meier method and Cox proportional hazards regression were used for unadjusted analysis of OS and PFS.

We explored three statistical methods that account for baseline imbalances between the treatment groups: multivariable regression analysis, 1:1 matching, and inverse probability of treatment weighting (IPTW). We applied the first two methods because these are most commonly used in the medical literature, while, at least in theory, IPTW may be preferred as a method in absence of a randomized trial, as explained below.

First, we used multivariable proportional hazards regression, with baseline variables as covariates in the model. The proportional hazards assumption was checked using a score test, which was valid for all variables, except possibly for the parameter that belonged to the patients with missing *BCL2*/*BCL6* status (p=0.055 in univariable analysis). For this category the relative hazard (compared with single-hit patients) decreased over time. However, this category is not of interest in itself, but was created so that patients with missing rearrangement status could be kept in the model. Therefore, the non-proportionality was ignored, assuming that this would be of little consequence for the hazard ratio (HR) of the treatment effect (Schemper, 1992). The HR from the multivariable regression is an adjusted HR and can be interpreted as the ratio of the hazards for two hypothetical patients whose only difference is that one was treated with R2CHOP, and the other with R-CHOP. The hazard ratio is non-collapsible, which means that the adjusted HR is not the same as the unadjusted HR, even if the treatment assignment (R2CHOP versus R-CHOP) would have been completely random. A consequence of this non-collapsibility property of the HR, is that the HR estimated by the multivariable regression here is not the same as the HR that would have been found in an unadjusted analysis of a randomized trial. The HR of the multivariable regression is called the 'conditional treatment effect', while the HR of a randomized trial is the 'average treatment effect' (ATE).

Second, we performed one-to-one matching. IPI score was used as the matching variable, because it is the most widely-used and validated prognostic score. Matching was performed without replacement (Austin, 2011), meaning that any patient treated with R-CHOP was only used once as a match for a patient treated with R2CHOP. Patients (from either group) were discarded if they had no available matches with the same IPI score (grouped as low, intermediate, or high), or if the set of their potential matches was depleted. By doing so, the IPI-risk distribution in the resulting matched dataset was artificially constructed to be equal to its distribution in the R2CHOP group. A limitation hereof is that the HR from this analysis estimates the treatment effect on patients treated with R2CHOP in this particular sample. This is called the 'average treatment effect on the treated' (ATT). The ATT

does not adequately estimate the ATE in the target population, which consists of all patients (i.e. both groups combined) in the present analysis.

Third, we performed inverse probability of treatment weighting (IPTW) using a propensity score. This is a method to estimate the ATE in the target population. These HRs are, therefore, most likely to reflect what would have been observed in an unadjusted randomized comparison, in contrast to multivariable analysis and matching. An advantage of IPTW over matching methods, is that IPTW uses the entire sample and does not need to remove patients for whom no match can be found.

We calculated a propensity score for being included in the HOVON-130 trial which was estimated from a logistic model based on the separate components of the IPI risk score, i.e., age, Ann Arbor stage, WHO performance score, number of extranodal localizations, serum LDH level. We additionally included sex and rearrangement status (single hit versus double/triple hit) in the model because these are known prognostic factors for overall survival. Patients with missing values were removed, except patients with unknown *BCL2*/*BCL6* rearrangement status; these were considered as a separate category. The propensity score was used for ITPW, where the IPTW weights were truncated at the 5-th and 95-th percentiles of their distribution. Standardized differences of baseline variables were calculated before and after weighting to assess the reduction of imbalances. The common support assumption was verified using the distribution of the propensity score across the treatment groups (figure 7.6 on page 112). To allow for some degree of misspecification of the model for the propensity score, we did a separate analysis using the augmented IPTW (AIPTW) estimator, which is a doubly-robust augmented IPTW estimator that combines inverse probability-of-censoring weighting (IPCW) and the g-formula (Ozenne et al., 2020). It is a doubly-robust estimator, which means that it is consistent if either the outcome model (for OS or for PFS), or the propensity model is correctly specified (but not if both are incorrect). Models for the treatment propensity, censoring, and outcome were specified using the same variables as used to estimate the propensity score (with actual treatment received also as variable in the latter two models). Thereby, we obtained absolute estimates of the ATE at 6 months and years 1 up to 5 since diagnosis. However, this method is not suited to estimate hazard ratios.

Analyses were performed in R, version 4.1.2 with package riskRegression version 2021.10.10.

7.3 Results

Patients

Of the 85 patients enrolled in the interventional R2CHOP cohort (HOVON-130 trial), 8 patients were ineligible for the present analysis (three because a *MYC* translocation could not be confirmed, and one because of transformed synchronous follicular lymphoma and four could not be identified in the NCR database). Data of 1171 (98%) of the 1200 DLBCL patients registered in the observational HOVON-900 cohort could be retrieved from the NCR. Of these, 56 (4.7%) fulfilled the eligibility criteria (i.e., eligible pathology review with *MYC*-R DLBCL, Ann Arbor stage II–IV and WHO performance status of 0–3) to serve as control in the present study (figure 7.1). Reasons for not being included in the interventional R2CHOP trial despite meeting its inclusion criteria were mainly logistic, e.g. the trial was not open in that center at that time, or the patient did not want to be referred or to participate.

Patients in the R2CHOP cohort received treatment between April 2015 and February 2018, and patients in the R-CHOP cohort between August 2015 to June 2019. Median follow-up was 4.16 years in patients treated with R2CHOP and 3.65 years in patients treated with R-CHOP ($p = 0.87$). The median time between diagnosis and start of treatment was 19 days (range 0–69 days) in the R2CHOP group and 15 days (range 5–84 days) in the R-CHOP group ($p = 0.317$).

Various baseline characteristics were imbalanced between the cohorts (table 7.1). Patients treated with R2CHOP were younger than patients treated with R-CHOP (median age 63 versus 70 years, $p = 0.018$), were more likely to have a lower WHO performance score ($p = 0.013$) and, as a consequence, had more often an intermediate IPI score (i.e. less often a low IPI score and less often a high IPI score, $p = 0.004$). There was no statistical proof that the distribution of sex, Ann Arbor stage, LDH levels, and rearrangement status were different between the cohorts, but there were numerical differences. For example, the R2CHOP group consisted of 18/77 SH patients (23.4%), 51/77DH patients (66.2%), and in 8 patients (10.4%) *BCL2* and *BCL6* status were both missing. In the R-CHOP cohort20/56 patients (35.7%) were SH, 26/56 patients (46.4%) were DH and in 10/56 patients (17.9%) *BCL2* and *BCL6* status were missing.

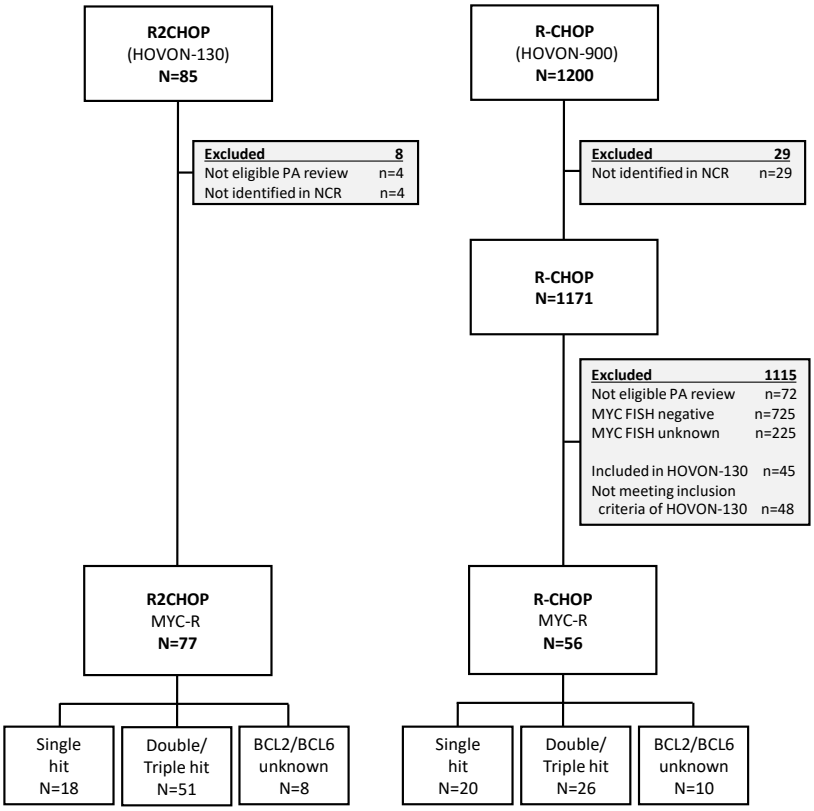


Figure 7.1: Flow chart of the patients included in the HOVON-130 and HOVON-900 for the current comparison

Table 7.1: Baseline characteristics by treatment group

	R-CHOP 56	R2CHOP 77	Total 133	*p*-value
Age at incidence (years)				0.018[1]
Median	70	63	66	
IQR	57–75	54–72	56–73	
Range	29–88	28–82	28–88	
(Missing)	0	0	0	

(Continued on next page)

Table 7.1 – continued from previous page

	R-CHOP 56	R2CHOP 77	Total 133	*p*-value
Sex				0.27[2]
Male	34 (60.7%)	54 (70.1%)	88 (66.2%)	
Female	22 (39.3%)	23 (29.9%)	45 (33.8%)	
(Missing)	0	0	0	
Ann Arbor stage				0.17[3]
2	12 (21.4%)	10 (13.0%)	22 (16.5%)	
3	12 (21.4%)	11 (14.3%)	23 (17.3%)	
4	32 (57.1%)	56 (72.7%)	88 (66.2%)	
(Missing)	0	0	0	
WHO performance score				0.013[3]
0	22 (41.5%)	47 (61.0%)	69 (53.1%)	
1	16 (30.2%)	24 (31.2%)	40 (30.8%)	
2	10 (18.9%)	5 (6.5%)	15 (11.5%)	
3	5 (9.4%)	1 (1.3%)	6 (4.6%)	
(Missing)	3	0	3	
WHO PS (grouped)				0.006[3]
0	22 (41.5%)	47 (61.0%)	69 (53.1%)	
1	16 (30.2%)	24 (31.2%)	40 (30.8%)	
2 or 3	15 (28.3%)	6 (7.8%)	21 (16.2%)	
(Missing)	3	0	3	
LDH				0.69[2]
Within normal range	16 (28.6%)	19 (25.0%)	35 (26.5%)	
Elevated	40 (71.4%)	57 (75.0%)	97 (73.5%)	
(Missing)	0	1	1	
Extranodal localizations				0.30[3]
None	12 (21.4%)	23 (29.9%)	35 (26.3%)	
1	22 (39.3%)	21 (27.3%)	43 (32.3%)	
2 or more	22 (39.3%)	33 (42.9%)	55 (41.4%)	
(Missing)	0	0	0	
IPI Risk Group				0.013[3]
Low	12 (21.8%)	9 (11.8%)	21 (16.0%)	
Low-intermediate	8 (14.5%)	22 (28.9%)	30 (22.9%)	
High-intermediate	13 (23.6%)	29 (38.2%)	42 (32.1%)	
High	22 (40.0%)	16 (21.1%)	38 (29.0%)	
(Missing)	1	1	2	
IPI Risk (3 Groups)				0.004[3]
Low	12 (21.8%)	9 (11.7%)	21 (15.9%)	
Intermediate	21 (38.2%)	52 (67.5%)	73 (55.3%)	
High	22 (40.0%)	16 (20.8%)	38 (28.8%)	
(Missing)	1	0	1	

(Continued on next page)

Table 7.1 – continued from previous page

	R-CHOP 56	R2CHOP 77	Total 133	p-value
Rearrangement				0.08[2]
Single hit	20 (35.7%)	18 (23.4%)	38 (28.6%)	
Double/triple hit	26 (46.4%)	51 (66.2%)	77 (57.9%)	
(Missing *BCL2*/*BCL6*)	10 (17.9%)	8 (10.4%)	18 (13.5%)	
Time to start treatment (days)				0.317[1]
Median	15.0	19.0	17.0	
IQR	10.8–23.8	11.0–26.0	11.0–26.0	
Range	5.0–84.0	0.0–69.0	0.0–84.0	
(Missing)	0	0	0	
Response				0.56[3]
Complete remission	37 (69.8%)	62 (80.5%)	99 (76.2%)	
Partial remission	11 (20.8%)	11 (14.3%)	22 (16.9%)	
Stable disease	1 (1.9%)	1 (1.3%)	2 (1.5%)	
Progressive disease	4 (7.5%)	3 (3.9%)	7 (5.4%)	
(Missing)	3	0	3	
Response (grouped)				0.21[2]
Complete remission	37 (69.8%)	62 (80.5%)	99 (76.2%)	
No complete remission	16 (30.2%)	15 (19.5%)	31 (23.8%)	
(Missing)	3	0	3	

1) Kruskal-Wallis rank sum test
2) Fisher's exact test
3) Trend test for ordinal variables

Overall Survival

The unadjusted OS of the patients treated with R2CHOP was significantly longer than in the R-CHOP cohort with a hazard ratio (HR) of 0.54 (95% CI 0.31–0.94, $p = 0.031$; figure 7.2A). To reduce bias resulting from baseline imbalances between the cohorts, we applied the three statistical methods described in the methods section.

First, in multivariable analysis, adjusting for the variables age at diagnosis, Ann Arbor stage, number of extranodal localizations, LDH, WHO performance status, and rearrangement status (table 7.2), yielded a comparable HR of 0.53 (95% CI 0.27–1.04, $p = 0.065$). Second, we performed an analysis of the patients who were matched on IPI score. For this analysis, 46 pairs could be analyzed (table 7.3) and an identical HR of 0.53 (95% CI 0.28–1.03, $p = 0.061$) was found. Third, estimation of the treatment effect of R2CHOP over R-CHOP on the total cohort by means of IPTW resulted in a HR of 0.59 (95% CI of 0.32–1.10, $p = 0.10$), and the absolute estimates using the doubly-robust method are shown in figure 7.2B. Assessments

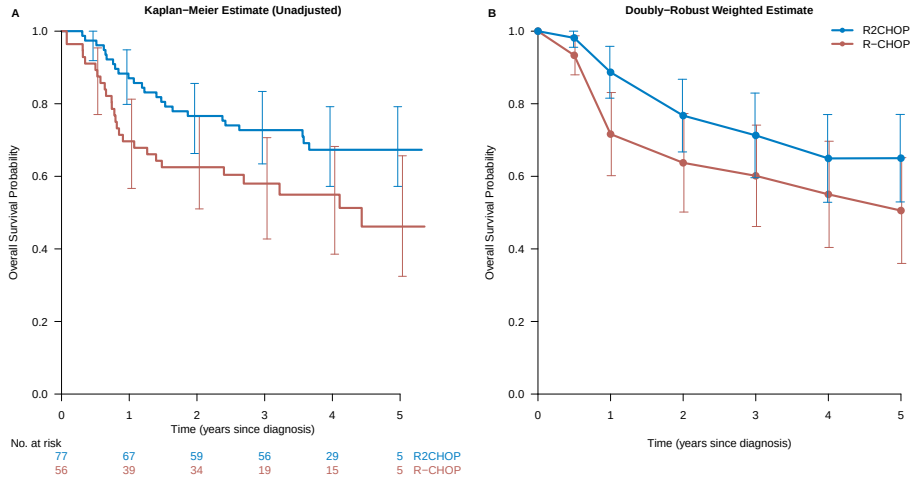


Figure 7.2: Overall survival analysis in *MYC*-R patients treated with R2CHOP in blue versus R-CHOP in red in (A) an unadjusted comparison of the overall survival by treatment and (B) doubly-robust analysis using AIPTW with IPCW estimate of overall survival

of the common support assumption and the reduction of imbalance are presented in the supplementary data (table 7.5 on page 111 and figure 7.6 on page 112).

Table 7.2: Multivariable Cox proportional-hazards regression of overall survival

	HR	95% CI	p-value
Treatment			
R-CHOP	1		
R2CHOP	0.53	(0.27 – 1.04)	0.065
Age at incidence			
(per year)	1.03	(1.00 – 1.06)	0.051
Ann Arbor stage			
2	1		
3	0.45	(0.14 – 1.47)	0.19
4	0.88	(0.29 – 2.62)	0.82
Extranodal localizations			
None	1		
1	0.57	(0.24 – 1.34)	0.20
2 or more	0.67	(0.29 – 1.56)	0.35
LDH			

(Continued on next page)

Table 7.2 – continued from previous page

	HR	95% CI	p-value
Within normal range	1		
Elevated	3.65	(1.31 – 10.17)	0.013
WHO PS (grouped)			
0	1		
1	1.32	(0.63 – 2.76)	0.45
2 or 3	2.05	(0.92 – 4.60)	0.08
Rearrangement			
Single hit	1		
Double/triple hit	1.02	(0.52 – 2.01)	0.95
Missing *BCL2*/*BCL6*	0.47	(0.15 – 1.49)	0.20

129 subjects, 48 events, 11 degrees of freedom
(4 subjects deleted due to missing values)

Table 7.3: Baseline characteristics in the IPI-matched set

	R-CHOP 46	R2CHOP 46	Total 92	p-value
Age at incidence (years)				0.11[1]
Median	70	65	68	
IQR	57–76	58–72	57–75	
Range	29–88	28–82	28–88	
(Missing)	0	0	0	
Sex				0.83[2]
Male	28 (60.9%)	30 (65.2%)	58 (63.0%)	
Female	18 (39.1%)	16 (34.8%)	34 (37.0%)	
(Missing)	0	0	0	
Ann Arbor stage				0.26[3]
2	9 (19.6%)	10 (21.7%)	19 (20.7%)	
3	11 (23.9%)	5 (10.9%)	16 (17.4%)	
4	26 (56.5%)	31 (67.4%)	57 (62.0%)	
(Missing)	0	0	0	
WHO performance score				0.17[3]
0	19 (43.2%)	30 (65.2%)	49 (54.4%)	
1	14 (31.8%)	11 (23.9%)	25 (27.8%)	
2	8 (18.2%)	4 (8.7%)	12 (13.3%)	
3	3 (6.8%)	1 (2.2%)	4 (4.4%)	
(Missing)	2	0	2	

(Continued on next page)

Table 7.3 – continued from previous page

	R-CHOP 46	R2CHOP 46	Total 92	*p*-value
WHO PS (grouped)				0.08[3]
0	19 (43.2%)	30 (65.2%)	49 (54.4%)	
1	14 (31.8%)	11 (23.9%)	25 (27.8%)	
2 or 3	11 (25.0%)	5 (10.9%)	16 (17.8%)	
(Missing)	2	0	2	
LDH				1.00[2]
Within normal range	13 (28.3%)	13 (28.3%)	26 (28.3%)	
Elevated	33 (71.7%)	33 (71.7%)	66 (71.7%)	
(Missing)	0	0	0	
Extranodal localizations				0.54[3]
None	11 (23.9%)	13 (28.3%)	24 (26.1%)	
1	17 (37.0%)	12 (26.1%)	29 (31.5%)	
2 or more	18 (39.1%)	21 (45.7%)	39 (42.4%)	
(Missing)	0	0	0	
IPI Risk Group				0.84[3]
Low	9 (19.6%)	9 (19.6%)	18 (19.6%)	
Low-intermediate	8 (17.4%)	11 (23.9%)	19 (20.7%)	
High-intermediate	13 (28.3%)	10 (21.7%)	23 (25.0%)	
High	16 (34.8%)	16 (34.8%)	32 (34.8%)	
(Missing)	0	0	0	
IPI Risk (3 Groups)				1.00[3]
Low	9 (19.6%)	9 (19.6%)	18 (19.6%)	
Intermediate	21 (45.7%)	21 (45.7%)	42 (45.7%)	
High	16 (34.8%)	16 (34.8%)	32 (34.8%)	
(Missing)	0	0	0	
Rearrangement				0.11[2]
Single hit	14 (30.4%)	15 (32.6%)	29 (31.5%)	
Double/triple hit	22 (47.8%)	28 (60.9%)	50 (54.3%)	
(Missing *BCL2*/*BCL6*)	10 (21.7%)	3 (6.5%)	13 (14.1%)	
Days before start treatment				0.07[1]
Median	15.5	21.5	19.0	
IQR	12.0–25.2	15.2–27.8	13.0–27.0	
Range	5.0–84.0	2.0–69.0	2.0–84.0	
(Missing)	0	0	0	

Response			0.35[3]
Complete remission	29 (67.4%)	38 (82.6%)	67 (75.3%)
Partial remission	10 (23.3%)	6 (13.0%)	16 (18.0%)
Stable disease	1 (2.3%)	0 (0.0%)	1 (1.1%)
Progressive disease	3 (7.0%)	2 (4.3%)	5 (5.6%)

1) Kruskal-Wallis rank sum test
2) Fisher's exact test
3) Trend test for ordinal variables

Progression-Free Survival

The unadjusted HR of PFS was 0.60 (95% CI 0.36–0.99, $p = 0.045$) in favor of R2CHOP (figure 7.3A). We analyzed the PFS using the same methods as for OS. Multivariable analysis resulted in a comparable HR of 0.59 (95% CI 0.33–1.08, $p = 0.085$, table 7.6 on page 112). In the set matched on IPI score a HR of 0.53 (95% CI 0.29–0.97, $p = 0.039$) was found. In the weighted analysis (ITPW), the HR was 0.60 (95% CI 0.32–1.12, $p = 0.11$) (figure 7.3B).

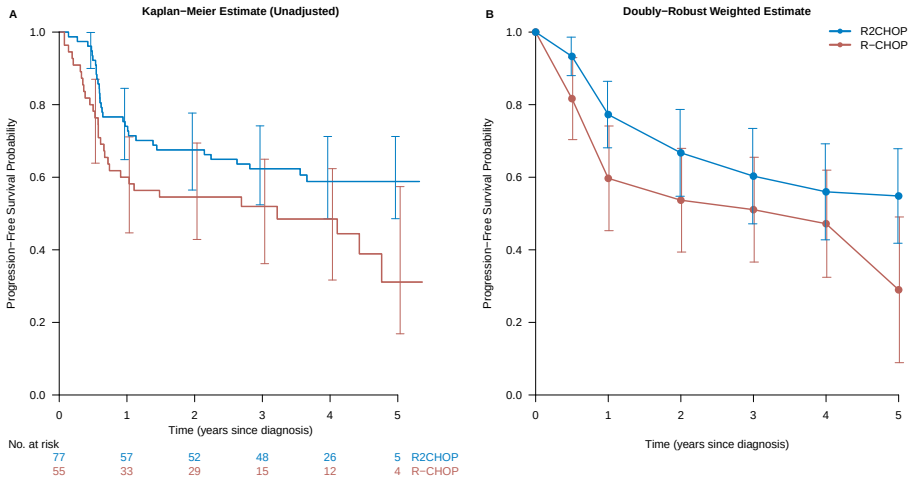


Figure 7.3: Progression-free survival analysis in *MYC*-R patients treated with R2CHOP in blue versus R-CHOP in red in (A) an unadjusted comparison of the overall survival by treatment and (B) doubly-robust analysis using AIPTW with IPCW estimate of progression-free survival

Overall survival and progression-free survival by rearrangement status

Because rearrangement status (SH or DH/TH) is known to be of prognostic importance for survival (although it was not statistically significant in our dataset, table 7.2), we did a subgroup analysis. Without any covariate adjustment, SH patients treated with R2CHOP tended

to have a longer OS than SH patients treated with R-CHOP with a HR of 0.34 (95% CI 0.10–1.10, $p = 0.072$) (figure 7.4A). Similarly, in the DH/TH subgroup, patients treated with R2CHOP tended to have a longer OS with a HR of 0.57 (95% CI of 0.28–1.13, $p = 0.11$, figure 7.4B). For PFS, the HRs were 0.66 in the SH subgroup (95% CI of 0.25–1.79, $p = 0.42$, figure 7.5A) and 0.48 in the DH/TH subgroup (95% CI of 0.26–0.90, $p = 0.022$, figure 7.5B). There were baseline imbalances within the subgroups (table 7.4 on 110), but we were unable to adjust properly due to low patient numbers.

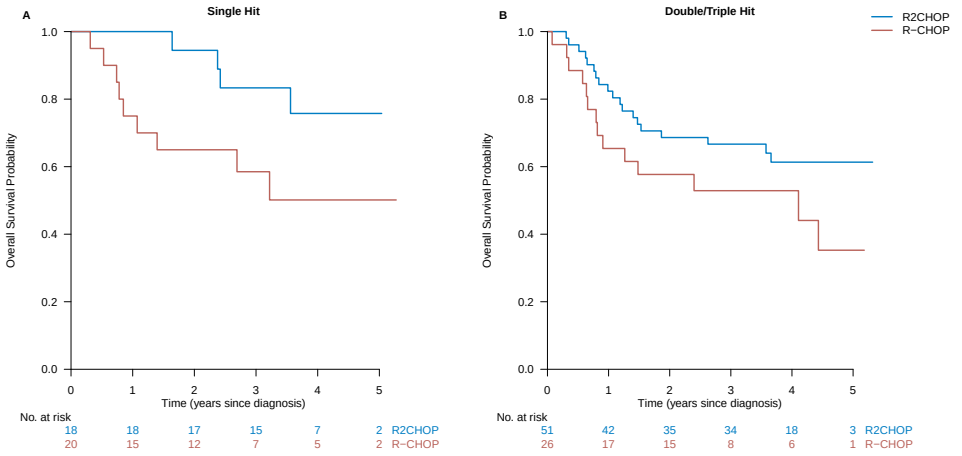


Figure 7.4: *MYC*-R patients treated with R2CHOP in blue versus R-CHOP in red in an unadjusted comparison depicted for (A) single hit patients and (B) double/triple hit patients

7.4 Discussion

To date, no published randomized trials were able to demonstrate improvements in overall survival over induction treatment with R-CHOP for patients with *MYC*-R DLBCL. Here, we present a comparison of addition of lenalidomide to R-CHOP (R2CHOP) versus R-CHOP as first-line treatment for newly diagnosed *MYC*-R DLBCL. We used long-term follow-up data of patients treated with R2CHOP in the single-arm phase-II HOVON-130 trial (Chamuleau et al., 2020). The analysis was extended by adding a cohort of patients who were treated with R-CHOP and met the inclusion criteria of the study, but were either not invited for logistic reasons, or who declined to participate in the trial. Because the two treatment regimens in this analysis were not randomized, any direct comparison is subject to treatment-selection bias due to systematic differences between the characteristics of the patients in the two groups. Therefore, we used three statistical methods (multivariable analysis, 1:1 matching and weighting using the propensity score) to reduce treatment-selection bias. These consistently showed improved survival after R2CHOP with HRs of approximately 0.59 for OS and 0.60 for PFS.

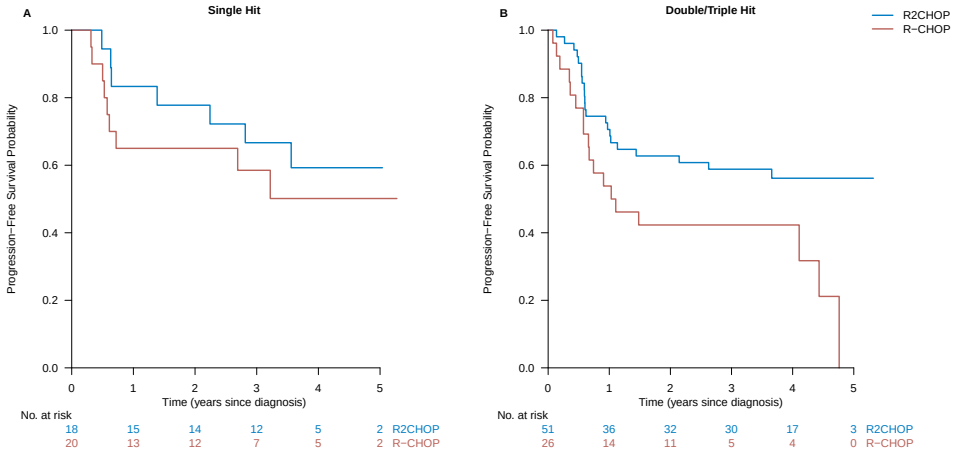


Figure 7.5: *MYC*-R patients treated with R2CHOP in blue versus R-CHOP in red in an unadjusted comparison depicted for (A) single hit patients and (B) double/triple hit patients

The most important limitation of weighting using the propensity score is that the propensity score has to be estimated using a statistical model, and it is impossible to verify whether this model was correctly specified (Austin & Stuart, 2015). Therefore, we used a doubly-robust method to circumvent this limitation. However, only a large randomized controlled trial (RCT) can balance observed as well as unobserved characteristics. Of the three statistical methods that we used, in theory weighting using the propensity score most closely resembles the result of an RCT, because it has the advantage that it estimates the average treatment effect on the entire sample. Yet, despite the relatively small sample size, the stability of the HRs for OS and PFS across the three methods lends credibility to the conclusion of a survival benefit for *MYC*-R patients treated with R2CHOP.

In the subgroups determined by rearrangement status, we performed unadjusted OS and PFS analysis, because we were not able to adjust for baseline imbalances because of the low number of events within the subgroups. These analyses can therefore not be interpreted as estimations of the effect of lenalidomide. Therefore, our findings on the survival benefit of R2CHOP apply only to all *MYC*-R patients. However, combined with the results of the multivariable analysis of the total cohort, they can be interpreted as an indication that the treatment effect is consistent across the subgroups. Whether patients in one subgroup benefit more from R2CHOP than another, needs to be investigated in future, larger studies.

We previously showed in the primary end-point analysis of the HOVON-130 study that the addition of lenalidomide to R-CHOP is well-tolerated and has limited and manageable adverse effects. There were no treatment-related deaths, and fewer grade 3 infections than expected for the more intensive immunochemotherapy regimen DA-EPOCH-R (Chamuleau et al., 2020). R2CHOP can be fully given on an outpatient basis, which is a major advantage to patients' well-being, and leads to a reduction in the costs of hospital admission days. Fur-

thermore, the fewer (serious) adverse effects that need to be treated, and the sharp drop in the price of lenalidomide in the EU since March 2022 after the patent expiration on lenalidomide, all result in making R2CHOP treatment likely to be cost-effective.

For current clinical practice, our results support further exploration of the addition of lenalidomide to R-CHOP as a valuable first-line treatment option for *MYC*-R DLBCL. Future studies may explore whether adding next-generation cereblon targeting antigens to R-CHOP can further increase the effects we observed here.

7.5 Supplementary Material

Supplementary methods

Missing values

The IPI score was calculated using the prognostic factors that it consists of (Age, Ann Arbor stage, WHO performance status, serum LDH level, and number of extra-nodal sites). If a missing value in these prognostic factors meant that the IPI score could not be calculated, it was set to missing. However, if for particular patients the presence of the risk factor would not change the score, the score was not considered missing. For example, if the performance status was missing but the patient had none of the other risk factors, the IPI risk group was set to low. Patients with missing IPI risk category were removed from the IPI-matched analysis.

The same principle was used for the categorization of the rearrangement status. This variable was constructed using the *BCL2*/*BCL6* information. Rearrangement status was considered missing if *BCL2* or *BCL6* measurement on FISH was missing, except if it could be determined for certain despite the missing values. For example, a *MYC*-R patient with a *BCL2* rearrangement a but *BCL6* missing, was counted in the double/triple hit category. Missing rearrangement status was used as a separate category in the propensity score model and multivariable proportional hazards regression. In the subgroup analysis of rearrangement status however, patients whose category (single hit, double/triple hit) could not unambiguously be determined were removed from the analysis.

Patients with missing values in other variables that were needed in the multivariable model or to calculate the propensity score, were removed from these particular analyses.

Table 7.4: Baseline Characteristics by Rearrangement Subgroup

	Single Hit		Double/Triple Hit	
	R-CHOP 20	R2CHOP 18	R-CHOP 26	R2CHOP 51
Age				
Median	67	62	72	65
IQR	58–72	57–70	59–75	54–72
Range	29–88	30–77	38–84	28–82
(Missing)	0	0	0	0
Sex				
Male	14 (70.0%)	14 (77.8%)	16 (61.5%)	34 (66.7%)
Female	6 (30.0%)	4 (22.2%)	10 (38.5%)	17 (33.3%)
(Missing)	0	0	0	0
Ann Arbor stage				
2	3 (15.0%)	4 (22.2%)	7 (26.9%)	6 (11.8%)
3	5 (25.0%)	1 (5.6%)	7 (26.9%)	8 (15.7%)
4	12 (60.0%)	13 (72.2%)	12 (46.2%)	37 (72.5%)
(Missing)	0	0	0	0
WHPO PS				
0	8 (42.1%)	13 (72.2%)	11 (45.8%)	30 (58.8%)
1	7 (36.8%)	3 (16.7%)	5 (20.8%)	17 (33.3%)
2	3 (15.8%)	2 (11.1%)	4 (16.7%)	3 (5.9%)
3	1 (5.3%)	0 (0.0%)	4 (16.7%)	1 (2.0%)
(Missing)	1	0	2	0
WHO PS (grouped)				
0	8 (42.1%)	13 (72.2%)	11 (45.8%)	30 (58.8%)
1	7 (36.8%)	3 (16.7%)	5 (20.8%)	17 (33.3%)
2 or 3	4 (21.1%)	2 (11.1%)	8 (33.3%)	4 (7.8%)
(Missing)	1	0	2	0
LDH				
Within normal range	8 (40.0%)	7 (41.2%)	6 (23.1%)	10 (19.6%)
Elevated	12 (60.0%)	10 (58.8%)	20 (76.9%)	41 (80.4%)
(Missing)	0	1	0	0
Extranodal localizations				
None	4 (20.0%)	3 (16.7%)	7 (26.9%)	17 (33.3%)
1	10 (50.0%)	9 (50.0%)	11 (42.3%)	10 (19.6%)
2 or more	6 (30.0%)	6 (33.3%)	8 (30.8%)	24 (47.1%)
(Missing)	0	0	0	0

(Continued on next page)

Table 7.4 – continued from previous page

	Single Hit		Double/Triple Hit	
	R-CHOP 20	R2CHOP 18	R-CHOP 26	R2CHOP 51
IPI Risk Group				
Low	5 (26.3%)	5 (29.4%)	6 (23.1%)	4 (7.8%)
Low-intermediate	4 (21.1%)	4 (23.5%)	3 (11.5%)	15 (29.4%)
High-intermediate	4 (21.1%)	4 (23.5%)	6 (23.1%)	20 (39.2%)
High	6 (31.6%)	4 (23.5%)	11 (42.3%)	12 (23.5%)
(Missing)	1	1	0	0
IPI Risk (3 Groups)				
Low	5 (26.3%)	5 (27.8%)	6 (23.1%)	4 (7.8%)
Intermediate	8 (42.1%)	9 (50.0%)	9 (34.6%)	35 (68.6%)
High	6 (31.6%)	4 (22.2%)	11 (42.3%)	12 (23.5%)
(Missing)	1	0	0	0
Response				
Complete remission	13 (68.4%)	16 (88.9%)	17 (68.0%)	40 (78.4%)
Partial remission	4 (21.1%)	1 (5.6%)	5 (20.0%)	8 (15.7%)
Stable disease	1 (5.3%)	0 (0.0%)	0 (0.0%)	1 (2.0%)
Progressive disease	1 (5.3%)	1 (5.6%)	3 (12.0%)	2 (3.9%)
(Missing)	1	0	1	0

Table 7.5: Standardized difference between the groups before and after weighting

Variable (or category)	Standardized difference before weighting	Standardized difference after weighting
Age	0.3711	0.0617
sex Female	0.1578	0.0034
Ann Arbor stage 3	0.2521	0.1017
Ann Arbor stage 4	-0.4049	-0.1408
WHO PS 1	-0.0298	0.0059
WHO PS 2 or 3	0.5700	0.1047
LDH Elevated	-0.1160	-0.0322
1 Extranodal localization	0.2864	0.0756
2 Or more extranodal localizations	-0.1148	-0.0802
Double/triple hit	-0.4497	-0.0645
Missing *BCL2*/*BCL6* status	0.2406	0.0584

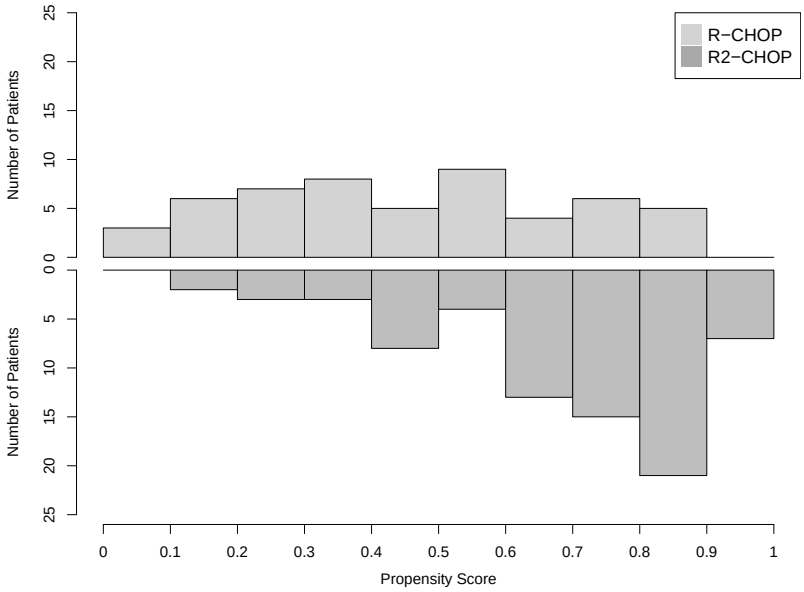


Figure 7.6: Histograms of the distribution of the propensity scores per treatment group

Table 7.6: Multivariable Cox proportional-hazards regression of progression-free survival

	HR	95% CI	p-value
Treatment			
R-CHOP	1		
R2CHOP	0.59	(0.33 – 1.08)	0.085
Age at incidence			
(per year)	1.01	(0.99 – 1.03)	0.53
Ann Arbor stage			
2	1		
3	0.58	(0.21 – 1.57)	0.29
4	0.93	(0.36 – 2.42)	0.88
Extranodal localizations			
None	1		
1	0.73	(0.34 – 1.56)	0.41
2 or more	0.61	(0.28 – 1.36)	0.23
LDH			
Within normal range	1		

(Continued on next page)

Table 7.6 – continued from previous page

	HR	95% CI	p-value
Elevated	2.88	(1.26 – 6.58)	0.012
WHO PS (grouped)			
0	1		
1	1.52	(0.78 – 2.95)	0.22
2 or 3	1.86	(0.88 – 3.92)	0.10
Rearrangement			
Single hit	1		
Double/triple hit	1.04	(0.56 – 1.93)	0.90
Missing *BCL2/BCL6*	0.49	(0.17 – 1.38)	0.17

129 subjects, 48 events, 11 degrees of freedom
(4 subjects deleted due to missing values)

Acknowledgements

The authors thank all study investigators and coordinators of the participating site of the HOVON-130 and HOVON-900 studies, the data-managers of the Netherlands Cancer Registry who helped retrieve the data, in particular Henrike Bretveld, and the HOVON Data Center in Rotterdam.

Celgene provided financial support of the HOVON-130 trial (004414). Dutch Cancer Society (KWF) provided financial support of the HOVON-900 (VUMC 2013-6269) and HOVON-130 (EMCR 2014-7436).

CHAPTER 8

Validation of a Treatment-Selection Rule for Patients with Advanced-Stage Ovarian Cancer

This chapter was published as:
R. van de Vrie*, E. van Werkhoven*, J. D. Asseler, M. J. Rutten, H. S. van Meurs, D. E. Werter, A. H. Zwinderman, M. R. Buist, C. A. R. Lok, P. M. M. Bossuyt, G. G. Kenter, and P. Tajik.
Validation of a treatment-selection rule for patients with advanced stage ovarian cancer.
European Journal of Gynaecological Oncology, 41(6):1023–1030, 2020.
*) Equally contributing

Abstract

Purpose of investigation: To externally validate the rule of Van Meurs et al. for selecting patients with advanced epithelial ovarian cancer for treatment with primary surgery or neoadjuvant chemotherapy (NACT).

Materials and Methods: We analysed a historical cohort of 900 consecutive patients with FIGO stage IIIC/IV ovarian cancer treated for advanced stage epithelial ovarian cancer at the Centre of Gynaecologic Oncology Amsterdam between 1998 and 2012. To externally validate the treatment-selection rule of Van Meurs et al. four groups were defined based on metastatic tumour size (smaller or larger than 45 mm) and FIGO stage (IIIC vs. IV). Within these groups, we compared survival outcomes of primary surgery and NACT.

Results: Differential treatment benefit in model-defined subgroups based on metastatic tumour size and FIGO stage was confirmed (interaction $p = 0.008$). Survival after primary surgery was significantly better compared to NACT plus interval debulking surgery for patients in FIGO stage IIIC ($p = 0.001$) or IV ($p = 0.028$) with metastases 45 mm, and those in FIGO stage IIIC with metastases $>$ 45 mm ($p = 0.011$). Survival was not significantly worse for FIGO stage IV patients with metastases $>$ 45 mm ($p = 0.094$). In patients with such large metastases, the location (omentum versus elsewhere in the body) was not prognostic ($p = 0.44$).

Conclusion: Our study has externally validated the treatment-selection rule first described by van Meurs et al. Primary surgery was shown to be superior for all patients except for the FIGO stage IV patients with a large metastatic tumour size ($>$ 45 mm), irrespective of localisation of the metastasis.

8.1 Introduction

Epithelial ovarian cancer is a leading cause of death from gynaecological malignancies worldwide. Most women are diagnosed with advanced disease (Federation of Gynaecology and Obstetrics (FIGO) stage III or IV) (Siegel et al., 2012), with a five-year overall-survival prognosis of 30% to 40% (Rutten et al., 2014; Van Meurs et al., 2013). Standard treatment consists of primary surgery followed by chemotherapy (Makar et al., 2016). An alternative treatment was widely implemented after publication of the European Organization for Research and Treatment of Cancer (EORTC) 55971 trial in 2010 (Vergote et al., 2010), which demonstrated non-inferiority of neoadjuvant chemotherapy (NACT) followed by interval debulking surgery and post-surgery chemotherapy. These results were confirmed in another randomized trial by Kehoe et al. (2015) and have been supported by other studies (Morrison et al., 2007; Vergote et al., 2011; Chi et al., 2011; Bristow et al., 2007; Vergote et al., 2016).

Despite this evidence, the choice between NACT and primary surgery remains controversial. NACT increases the chance of total removal of all macroscopic tumour and could lead to improved survival, since residual disease after debulking surgery is the most important prognostic factor for survival (Vergote et al., 1998). The guideline of the American Society of Clinical Oncology recommends NACT for women with a high risk profile and a low likelihood of cytoreduction to < 1 cm (Wright et al., 2016). Models to identify such patients have been developed based on patient characteristics, laparoscopy results, imaging features, or combinations of these (Rutten et al., 2015, 2017)

However, residual disease after debulking surgery is only a surrogate endpoint for survival and is difficult to predict (Rutten et al., 2015). To solve this problem, treatment-selection models have been developed that focus directly on survival, such as the prediction model developed by Van Meurs et al. (2013). In a secondary analysis of the EORTC 55971 trial, they demonstrated that FIGO stage and the size of the largest metastatic tumour are significantly associated with benefit from treatment. They selected ten baseline clinical and pathological characteristics as potential biomarkers. Using Subpopulation Treatment Effect Pattern Plots (STEPP), they considered biomarkers with a statistically significant qualitative additive interaction with treatment as being potentially informative for treatment selection. Their study showed that the size of the largest metastatic tumour and the FIGO stage were significantly associated with magnitude of benefit from treatment. These were combined to create a multimarker treatment-selection rule: primary surgery was recommended for patients with FIGO stage IIIC and a largest metastatic tumour size ≤ 45 mm. Both primary surgery and NACT were feasible options for patients with FIGO stage IIIC with largest metastatic tumour size > 45 mm and for FIGO stage IV patients with largest metastatic tumour size ≤ 45 mm. NACT was recommended

for patients with FIGO stage IV disease with a largest metastatic tumour size > 45 mm.

This treatment-selection rule has not yet been externally validated, which prevents unconditional recommendation of its use. In the present study, we therefore performed such an external validation. In addition, we evaluated whether the localisation of the metastasis—within the omentum compared to a metastasis at another site—influences survival differentially.

8.2 Materials and Methods

We conducted a multicentre historical cohort study of consecutive patients who were treated for advanced-stage epithelial ovarian cancer at the Centre of Gynaecologic Oncology Amsterdam (Academic Medical Centre, Free University Medical Centre and Antoni van Leeuwenhoek) between 1998 and 2012. We included all women who had epithelial cancer of the ovary, tube or peritoneum classified as advanced stage based on FIGO criteria (stages IIIC or IV) and who underwent either primary debulking surgery or NACT with interval debulking surgery. To select only patients for whom it is more difficult to predict whether cytoreductive surgery will be successful, we excluded those with lower stage ovarian cancer (FIGO stage I–IIIB). To ensure a truly external validation (without overlap with the dataset that was used to develop the decision rule) patients who participated in the EORTC 55971 trial were excluded. Finally, to avoid bias from this additional treatment within the NACT group, we excluded patients who received HIPEC treatment as part of the OVHIPEC trial (van Driel et al., 2018).

Data were retrieved from medical files, including FIGO stage and largest metastatic tumour size in mm (including omental cake). In women who received primary surgery, the largest metastatic tumour size was evaluated by either the CT scan prior to surgery or the surgery report. For women who received NACT with interval surgery, the metastatic tumour size was evaluated on the CT scan made prior to chemotherapy treatment. If this was inconclusive, a sonography image prior to treatment was used. If no specific measurement for metastasis was found in the medical records, metastatic tumour size was considered missing.

As defined by Van Meurs et al. (2013), the following four subgroups were constructed:

1. FIGO stage IIIC and largest metastatic tumour size ≤ 45 mm;
2. FIGO stage IIIC and largest metastatic tumour size > 45 mm;
3. FIGO stage IV and largest metastatic tumour size ≤ 45 mm;
4. FIGO stage IV and largest metastatic tumour size > 45 mm.

Baseline characteristics were compared between treatment groups using Wilcoxon's rank sum test and Fisher's exact test or the χ^2-test. Cox propor-

tional hazards models of overall survival were used to analyse the treatment effect and the treatment-by-subgroup interaction. Subsequently, separate Cox proportional hazards models were used to analyse overall survival by treatment in the four subgroups. Missing data were imputed, i.e. they were replaced with substitute values that are based on a regression model. The imputation procedure was repeated 10 times using the multiple imputation method, which ensures that variation introduced by the imputation process is accounted for. The event indicator and the Nelson-Aalen estimator of the hazard were included in the imputation model.

After multiple imputation, predictions are usually calculated from the average coefficients (i.e. the log hazard ratios) across the imputed datasets, which are combined into a linear predictor. To ensure the validity of this method, each of the separate imputation sets must contain data from the same patients. However, this was not the case in our analysis for the models in the separate subgroups because the size of the largest metastasis was missing for some patients, and the subgroups to which such patients belonged could vary between multiple-imputation datasets. Therefore, survival curves per subgroup were obtained from the average of the linear predictors instead of the average of the coefficients. This was done as follows: for each patient occurring in a subgroup at least once, the mean linear predictor was calculated across the imputations for which that patient belonged to that subgroup. These linear predictors were then used to calculate the Breslow estimator of the survival curve of that subgroup.

In contrast to the study design of Van Meurs et al. (2013) we did not collect data from a randomized study but from patient records. To account for possible treatment-by-indication bias, baseline imbalances between the two treatment groups were compensated for by weighting patients in each model by the inverse probability of the treatment they actually received. The following variables were used to calculate the probability of receiving treatment: age, BMI, FIGO stage, size of metastasis, CA-125, albumin, grade, ASA, platelets, location of first debulking surgery, ascites (whether it was present and was more than or less than 500 ml) and period of treatment (before 2010 or after). The period was included to account for possible changes after the publication of the EORTC 55971 trial (Vergote et al., 2010). BRCA status of the majority of patients was unknown and could therefore not be included in the analyses. The weights were trimmed to the 12^{st} and 992^{th} percentile of their distribution. To assess the effect of the weighting on the imbalance between the treatment groups, standardized differences were calculated (details are provided in the appendix on page 129).

A sensitivity analysis was done with separate Cox proportional hazards models of overall survival by treatment in the four subgroups without imputation or weighting, using only patients with known metastatic tumour size. Another analysis was

done in the subgroup of patients with a large metastatic tumour (> 45 mm), in which survival was analysed by whether patients had a large metastasis within the omentum or a large metastasis anywhere else in the body. A p-value of < 0.05 was considered to indicate statistically significant differences. Statistical analyses were performed with the Statistical Package for the Social Sciences software package, version 23 for Windows (SPSS Inc., Chicago, Illinois, USA) and R 3.4.4 with package mice version 2.46.0.

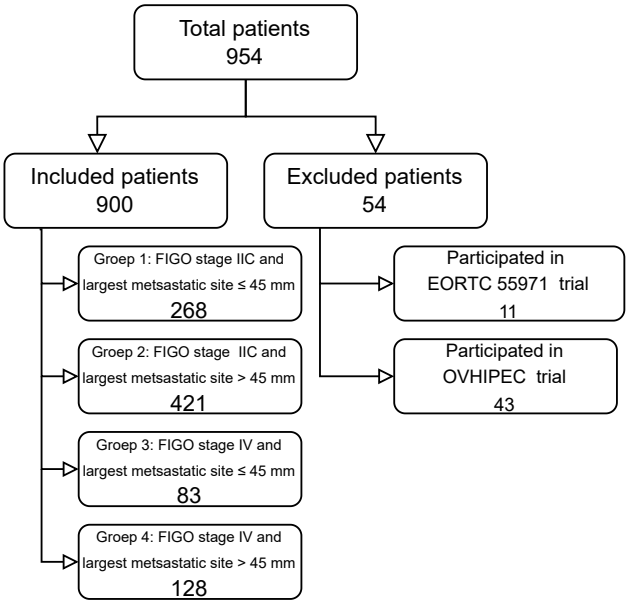


Figure 8.1: Flow diagram of included patients

8.3 Results

We identified 954 eligible patients in the study period, of which 54 were excluded because of participation in the EORTC 55971 trial ($n = 11$) or the OVHIPEC trial ($n = 43$, figure 8.1). Of the remaining 900 patients, 319 (35%) underwent primary surgery and 581 patients (65%) received NACT followed by interval surgery. Of all patients, 689 (77%) had clinical FIGO stage IIIC, and 211 (23%) had FIGO stage IV (Table 1). FIGO stage was available for all patients, but the size of the metastatic tumour was missing in 244 patients (27%).

Table 8.1: Patient characteristics

	Primary surgery 319	NACT followed by interval surgery 581	p-value
Age (years)			< 0.001
Median (IQR)	60 (51–68)	(56–70)	
Body Mass Index (BMI)			0.95
Median (IQR)	24.6 (21.9–27.8)	24.5 (21.8–27.8)	
(Missing)	129 (40%)	193 (33%)	
Post-menopausal			< 0.001
No	45 (17%)	38 (8%)	
Yes	217 (83%)	459 (92%)	
(Missing)	57 (18%)	84 (14%)	
WHO performance status			0.002
0	166 (60%)	220 (49%)	
1	98 (35%)	172 (38%)	
2	12 (4%)	49 (11%)	
3	2 (1%)	8 (2%)	
(Missing)	41 (13%)	132 (23%)	
Tumour grade by Silverberg			0.49
1	18 (7%)	19 (7%)	
2	60 (22%)	54 (18%)	
3	189 (71%)	219 (75%)	
(Missing)	52 (16%)	289 (50%)	
Histology			< 0.001
Serous	228 (71%)	462 (80%)	
Mucinous	14 (4%)	13 (2%)	
Endometroid	36 (11%)	15 (3%)	
Clear cell	13 (4%)	13 (2%)	
Undifferentiated	19 (6%)	74 (13%)	
Mixed	2 (1%)	2 ($< 1\%$)	
Other	7 (2%)	2 ($< 1\%$)	
Serum CA-125 before treatment (U/ml)			< 0.001
Median (IQR)	609 (218–1757)	963 (380–2413)	
(Missing)	13 (4%)	30 (5%)	
Amount of ascites before treatment			0.12
> 500 ml	179 (60%)	233 (54%)	
≤ 500 ml	118 (40%)	197 (46%)	
(Missing)	22 (7%)	151 (26%)	
Clinical stage			< 0.001
IIIC	288 (90%)	401 (69%)	
IV	31 (10%)	180 (31%)	
(Missing)	0 (0%)	0 (0%)	
Metastatic tumour size			0.91
≤ 45 mm	103 (41%)	160 (40%)	
> 45 mm	151 (59%)	242 (60%)	
(Missing)	65 (20%)	179 (31%)	

(Continued on next page)

Table 8.1 – continued from previous page

	Primary surgery 319	NACT followed by interval surgery 581	p-value
Subgroup			< 0.001
1. FIGO stage IIIC and largest metastatic tumour size ≤ 45 mm	94 (29%)	108 (19%)	
2. FIGO stage IIIC and largest metastatic tumour size > 45 mm	135 (42%)	172 (30%)	
3. FIGO stage IV and largest metastatic tumour size ≤ 45 mm	9 (3%)	52 (9%)	
4. FIGO stage IV and largest metastatic tumour size > 45 mm	16 (5%)	70 (12%)	
(Missing)	65 (20%)	179 (31%)	
Residual tumour after surgery			< 0.001
No residue	96 (30%)	269 (46%)	
< 1 cm residue	75 (23%)	231 (40%)	
> 1 cm residue	148 (46%)	80 (14%)	
(Missing)	0 (0%)	1 (< 1%)	

Some patient characteristics differed between the treatment groups. The NACT group was on average older ($p < 0.001$), with a corresponding higher percentage postmenopausal women ($p < 0.001$). World Health Organization (WHO) performance status was higher in the NACT group ($p = 0.002$),

and histology differed between the two treatment groups ($p < 0.001$). Significantly more patients with FIGO stage IIIC were present in the primary surgery group, and significantly more patients with FIGO stage IV were present in the NACT group ($p < 0.001$). Imbalance in treatment period was most prominent in subgroup 4 (FIGO stage IV and largest metastatic tumour size > 45 mm), where all 16 patients who received primary surgery were treated between 1997 and 2008. Serum CA-125 before treatment ($p < 0.001$) was unevenly distributed, with higher levels in the NACT group. After weighting, baseline characteristics of the treatment groups were more comparable (tables 8.3 and 8.4 on page 131). We also inspected the distribution of propensity scores per treatment group and found sufficient overlap (figure 8.5 on page 133).

After 10 multiple imputations, the average distribution of patients across the four subgroups was as follows: Group 1: 268 (30%) patients had FIGO stage IIIC and largest metastatic size ≤ 45 mm; Group 2: 421 patients (47%) had FIGO stage IIIC and largest metastatic size > 45 mm; Group 3: 83 patients (9%) had FIGO stage IV and largest metastatic size ≤ 45 mm; and Group 4: 128 patients (14%) had FIGO

stage IV and largest metastatic size > 45 mm (table 8.2).

Table 8.2: Number of patients per subgroup defined by clinical (FIGO) stage and the largest metastatic tumour size with the comparative outcome of primary surgery and neoadjuvant chemotherapy in terms of five-year survival for each subgroup

Subgroup	Patients*	5-year survival probability		NACT *vs.* primary surgery HR (95% CI)	p-value
		prim. surg.	NACT		
1. FIGO stage IIIC and largest metastatic size ≤ 45 mm	268 (30%)	42%	27%	1.51 (1.18–1.93)	0.001
2. FIGO stage IIIC and largest metastatic size > 45 mm	421 (47%)	29%	21%	1.25 (1.05–1.48)	0.011
3. FIGO stage IV and largest metastatic size ≤ 45 mm	83 (9%)	35%	17%	1.66 (1.06–2.59)	0.028
4. FIGO stage IV and largest metastatic size > 45 mm	128 (14%)	10%	17%	0.77 (0.57–1.05)	0.094
				p-value for interaction $= 0.008$	

*) Using average size of metastatic tumour across 10 multiple imputations

The median follow-up was 94 months; 38 patients' survival time was censored before five years of follow up. Overall survival was significantly worse in the group undergoing NACT compared to primary surgery (HR 1.31, 95% CI 1.17–1.45, $p < 0.0001$, figure 8.2).

In the subgroup of patients with FIGO stage IIIC and metastatic size ≤ 45 mm, the five-year survival was 27% in the NACT group, versus 42% in the primary surgery group (HR 1.51, 95% CI 1.1–81.93, $p = 0.001$). In patients with stage IV and ≤ 45 mm there was also a benefit from primary surgery (HR 1.66, 95% CI 1.06–2.59, $p = 0.028$) and, to a lesser extent, in the subgroup of patients with stage IIIC and metastatic size > 45 mm (HR 1.25, 95% CI 1.05–1.48, $p = 0.011$). In the subgroup of stage IV and metastatic size of > 45 mm, the effect appeared to be reversed, although the difference in survival was not significant (HR 0.77, 95% CI 0.57–1.05, $p = 0.094$). The predicted five-year survival in this subgroup was 10% with primary surgery and 17% with NACT (figure 8.3).

The heterogeneity of the treatment effect across the four subgroups was statistically significant (p for interaction = 0.008). In the subgroup of patients with a large metastatic tumour who received primary surgery, no significant survival difference was observed between patients with a large metastasis within the omentum and patients with a large metastasis anywhere else in the body (HR 0.81, 95% CI 0.47–

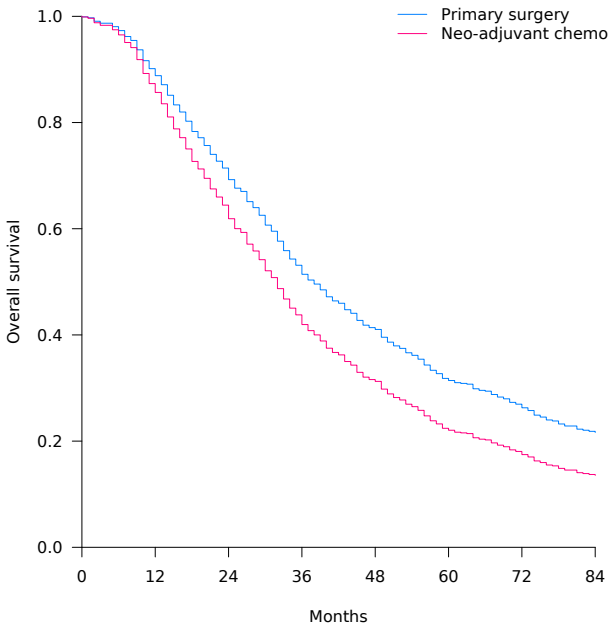


Figure 8.2: Overall survival by treatment strategy, HR 1.31 95%CI (1.17–1.45), $p < 0.0001$

1.40, $p = 0.44$, figure 8.4).

The sensitivity analysis without weighting in the 656 patients with known metastatic tumour size showed a significant benefit in the subgroup with FIGO stage IIIC and metastatic size 45 mm for primary surgery (HR 1.87, 95% CI 1.35–2.59, $p = 0.0001$) and the subgroup with FIGO IIIC and metastatic size > 45 mm (HR 1.37, 95% CI 1.07–1.75, $p = 0.013$), but not for the two FIGO IV subgroups (figure 8.6 on page 134).

8.4 Discussion

The aim of our study was to externally validate the treatment-selection rule by Van Meurs et al. (2013) for selecting the type of treatment of advanced stage ovarian cancer. We found that patients with FIGO stage IIIC or IV with a small metastatic tumour (≤ 45 mm), or with FIGO stage IIIC and a large metastatic tumour (> 45 mm) benefit most from primary surgery. No significant survival difference was found for patients with stage IV and a large metastatic tumour size (> 45 mm). This partly confirms the conclusions of Van Meurs et al. (2013) and is consistent

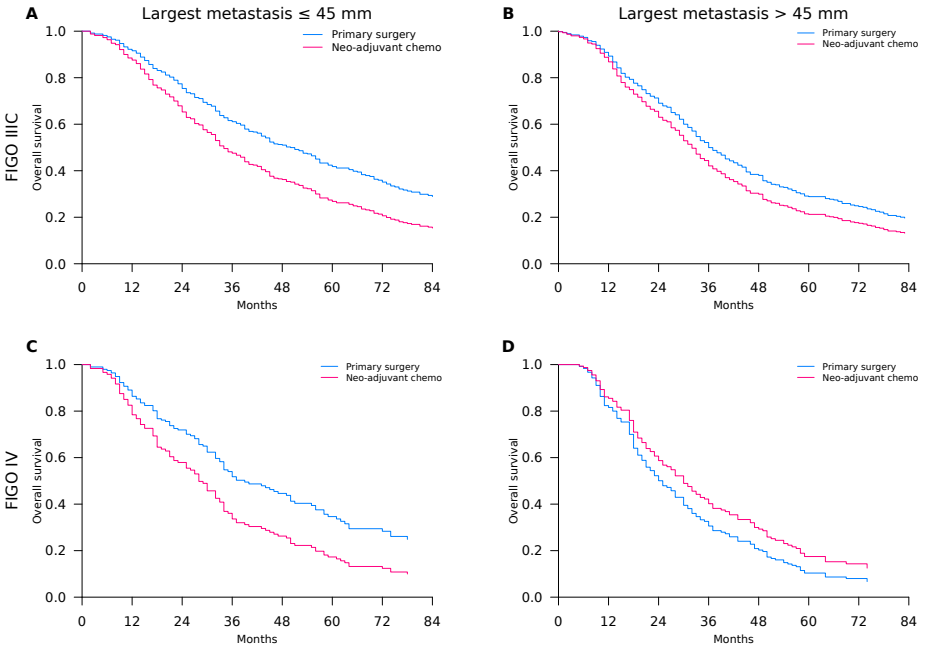


Figure 8.3: Overall survival by treatment in the four biomarker subgroups
(A) Subgroup 1 of patients with FIGO stage IIIC and metastaticsize ≤ 45 mm, HR 1.51 95% CI (1.18–1.93), $p = 0.001$.
(B) Subgroup 2 of patients with FIGO stage IIIC and metastatic size > 45 mm, HR 1.25 95% CI (1.05–1.48), $p = 0.011$.
(C) Subgroup 3 of patients with FIGO stage IV and metastatic size ≤ 45 mm, HR 1.66 95% CI (1.06–2.59), $p = 0.028$.
(D) Subgroup 4 of patients with FIGO stage IV and metastatic size > 45 mm, HR 0.77 95% CI (0.57–1.05), $p = 0.0944$.
The heterogeneity of the treatment effect across the four subgroups was statistically significant (p for interaction = 0.008)

with the results of Meyer et al. (2016). In both of these studies, primary surgery was associated with increased survival in stage IIIC, but not for stage IV disease.

We expected that a metastasis within the omentum, which can be resected relatively easily, would have a better prognosis compared to a metastasis anywhere else in the body. However, this difference was not seen in the subgroups of patients with a large metastatic tumour. This suggests that the extent of disease may be more important for the survival prognosis than the exact localisation of the metastasis.

The strength of our study lies in the large number of included patients and the com-

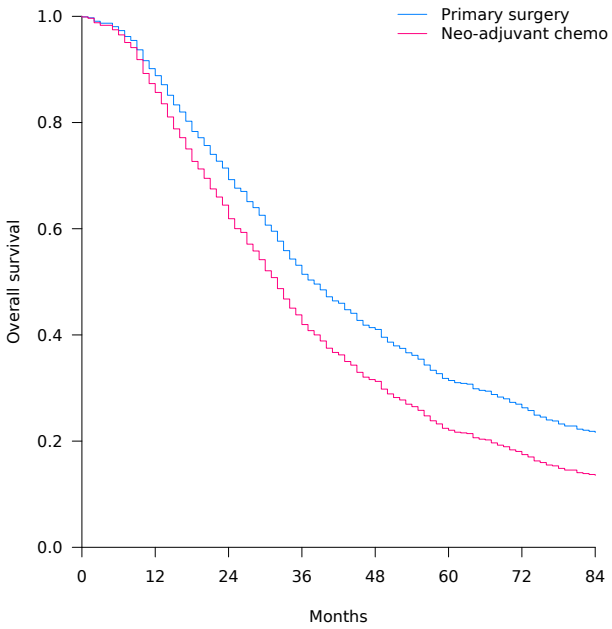


Figure 8.4: Subgroup analysis in patients with a large metastatic tumour size > 45 mm who received primary surgery by whether the large metastasis was within the omentum or anywhere else in the body. HR 0.81 $(0.47–1.40)$, $p = 0.44$

pleteness and duration of follow up. In this dataset of 900 patients, only 38 patients had times censored before five years. Generalizability was increased by combining data from three large institutes. These institutes were oncological centres, so all patients were treated by experienced gynaecological oncologists.

Several limitations should also be mentioned. As with all non-randomised studies, estimates from this study may suffer from selection bias. Firstly, only patients with a primary or interval surgery were included. This means that patients were not included if they started NACT but never underwent surgery, for example because of deterioration of their physical condition due to toxicity of neoadjuvant chemotherapy or tumour growth during NACT. Secondly, there were imbalances in the baseline characteristics between the two treatment groups: the significantly lower WHO performance status in the primary surgery group could have affected survival outcome. Moreover, primary surgery was the standard treatment in the Netherlands before 2010, and interval surgery was mostly given to patients who were deemed unfit for primary surgery. After 2010, the results of the EORTC 55971

trial (Vergote et al., 2010) influenced the selection for various treatment approaches, with a high rate of NACT. We attempted to correct for these imbalances by using inverse probability of treatment weighting, where the weights were calculated using a model for the probability of getting the actual treatment. This model included baseline characteristics and the period of treatment (before 2010 or after). Despite this correction we still observed a difference between the two treatments, even though non-inferiority was demonstrated by Vergote et al. (2010). This may indicate that there are unknown confounding factors that we could not take into account. This emphasizes the importance of randomized clinical trials.

Another disadvantage of the retrospective data collection in this study was the missing data. For example, metastatic tumour size was missing for 244 of the 900 patients (27%). This is not surprising because usually not all localisations of metastatic disease are measured separately during surgery. Multiple imputation was used for these and other variables with missing values. In a sensitivity analysis without weighting, using only patients with known metastatic tumour size, the hazard ratios did not differ very much, but were only statistically significant in the two FIGO IIIC subgroups. But even the non-missing values of the metastatic tumour size have a degree of uncertainty, because these were based on measures as they were reported in the past, with a risk of false classification and patients being classified in the wrong subgroup. BRCA status was missing for the majority of patients and could therefore not be imputed reliably. An imbalance in BRCA status could have influenced the result of our study because of the improved OS due to sensitivity for platinum and PARP inhibition. However, it is unlikely that this would have led to the impossibility of validating the treatment model, because also in BRCA mutation carriers, tumour size of metastasis and FIGO stage are likely to be prognostic.

Possibilities for further research include the development of various treatment-selection rules to predict residual tumour after debulking surgery, which are currently being developed and tested based on patient characteristics, laparoscopy results, imaging features, or combinations of these variables (Rutten et al., 2015, 2017). For example, Rutten et al. (2015) recommended adding an open laparoscopy to the standard diagnostic work-up, and Fagotti et al. (2008) described an easy-to-use prediction model for laparoscopy to guide treatment. Their model calculates a score from seven laparoscopic features. Laparoscopic assessment after standard diagnostic work-up could potentially be combined with our treatment-selection rule for the subgroups with uncertain results.

Many prediction models are being developed, but only a few have been evaluated with external data. Consequently, there is insufficient knowledge about their performance, which may explain why many of them are not used in clinical practice Collins et al. (2014).

8.5 Conclusions

Based on this analysis, we conclude that overall survival after primary surgery is better than after NACT followed by interval surgery in a subgroup of patients. We confirmed the treatment difference as described by Van Meurs et al. (2013) for patients with a small metastatic tumour size of ≤ 45 mm and FIGO stage IIIC or IV, and for FIGO stage IIIC with a large metastatic tumour size > 45 mm, with better survival after primary surgery. However, for patients with a large metastatic tumour size (> 45 mm) and FIGO stage IV, survival was not significantly different. Overall, the results of this external validation provide further support for the recommendations of Van Meurs et al. (2013) and can facilitate a more widespread use of this model for making recommendations about treatment options in women with advanced stage ovarian cancer.

Ethics approval and consent to participate

From all patients, written informed consent was provided to conduct retrospective research on clinical data. The study was conducted in accordance with the Decleration of Helsinki, due to the retrospective nature of the analyses approval by the Ethics Committee was not necessary.

8.6 Supplementary material

Effect of weighting on the imbalance between the treatment groups

In order to quantify the amount of imbalance between the two treatment groups, balance diagnostics were calculated. A commonly used diagnostic is the standardized difference (Austin, 2009). For continuous variables, the standardized difference is defined as

$$d = \frac{(\bar{x}_A - \bar{x}_B)}{\sqrt{\frac{s_A^2 + s_B^2}{2}}}, \tag{8.1}$$

which measures the difference in means in units of the pooled standard deviation. Here $\bar{x}_A$ and $\bar{x}_B$ are the sample means in the two treatment groups and s_A^2 and s_B^2 are the sample variances. This diagnostic is often used to see if after matching, baseline covariates are more equally distributed across the treatment groups.

The above formula was extended in order to apply the same idea to weighted data instead of matched data. The extension consists of incorporating the case weights into the formula, and taking into account that the sizes of the treatment groups need not be equal.

The weighted standard deviation was defined as

$$s_w = \sqrt{\frac{\sum_{i=1}^{n} w_i (x_i - \hat{\mu}^*)^2}{\mathrm{V}_1 - (\mathrm{V}_2/\mathrm{V}_1)}}, \tag{8.2}$$

where $\hat{\mu}^* = \frac{\sum_i w_i x_i}{\sum_i w_i}$ is the weighted arithmetic mean, and

$$\mathrm{V}_1 = \sum_{i=1}^{n} w_i \tag{8.3}$$

$$\mathrm{V}_2 = \sum_{i=1}^{n} w_i^2 \tag{8.4}$$

are the sums of the weights and the squared weights, respectively.

The pooled version of the weighted standard deviations $s_{w,A}$ in treatment A and $s_{w,B}$ in treatment group B was then defined as

$$s_{w,\text{pooled}} = \sqrt{\frac{(\mathrm{V}_{1,A} - \mathrm{V}_{2,A}/\mathrm{V}_{1,A}) s_{w,A}^2 + (\mathrm{V}_{1,B} - \mathrm{V}_{2,B}/\mathrm{V}_{1,B}) s_{w,B}^2}{(\mathrm{V}_{1,A} - \mathrm{V}_{2,A}/\mathrm{V}_{1,A}) + (\mathrm{V}_{1,B} - \mathrm{V}_{2,B}/\mathrm{V}_{1,B})}} \tag{8.5}$$

and the weighted standardized difference as

$$d_w = \frac{\hat{\mu}_A^* - \hat{\mu}_B^*}{s_{w,\text{pooled}}}. \tag{8.6}$$

For categorical variables, the standardized difference was calculated by coding one category as a 1 and the other as a zero. For categorical variables with three or more category levels, the difference was calculated separately for each category compared to the other categories combined. For example, the outcome of the surgery had categories 'no residual tumour', 'residual tumour < 1 cm', and 'residual tumour > 1 cm'. So for this variable three differences were calculated, each comparing one of the categories with the two other categories.

The results presented in table 8.3 are for observed observations only (no imputation done). A value of NA in this table indicates that there were no patients in that group for whom a weight could be calculated because of missing values. Results of the balance diagnostics across multiple imputations are presented in table 8.4.

Table 8.3: Standardized difference between the groups before and after weighting (complete observations)

	Standardized difference	
	without weighting	after weighting
Age	-0.3229	-0.0133
BMI	0.0335	0.0397
Postmenopausal	-0.3082	-0.2487
WHO performance status	-0.2787	-0.0307
Tumour grade by Silverberg	-0.0745	-0.0382
Histology:		
serous adenocarcinoma	-0.1908	0.0370
mucinous adenocarcinoma	0.1262	-0.1305
endometriod adenocarcinoma	0.3823	0.2531
clearcell adenocarcinoma	0.1097	-0.0741
undifferentiaded adenocarcinoma	-0.2238	-0.1214
mixed epithelial adenocarcinoma	0.0425	NA
other epithelial tumor	0.1865	NA
Serum CA-125 before treatment	-0.1379	-0.0233
Amount of ascites before treatment	-0.1228	0.0082
Clinical stage	-0.5165	-0.0388
Metastatic tumour size	-0.0153	0.0667
Residual tumour after surgery:		
no macroscopic residual tumour	-0.3355	-0.1904
residual tumour < 1 cm	-0.3487	-0.2763
residual tumour > 1 cm	0.8018	0.4846

Table 8.4: Standardized difference between the groups after weighting (averaged across multiple imputations)

	Mean standardized difference after weighting
Age	-0.0130
BMI	0.0214
Postmenopausal	-0.0672
WHO performance status	-0.0637
Tumour grade by Silverberg	0.0078
Histology:	
serous adenocarcinoma	-0.0663
mucinous adenocarcinoma	0.0567
endometriod adenocarcinoma	0.2887
clearcell adenocarcinoma	0.0091
undifferentiaded adenocarcinoma	-0.2738
mixed epithelial adenocarcinoma	0.0107
other epithelial tumor	0.1952
Serum CA-125 before treatment	0.0414
Amount of ascites before treatment	-0.0757
Clinical stage	-0.0857
Metastatic tumour size	-0.0157
Residual tumour after surgery:	
no macroscopic residual tumour	-0.3985
residual tumour < 1 cm	-0.3642
residual tumour > 1 cm	0.8472

Histograms of the Propensity Scores

The distribution of the propensity scores across the two treatment goups was inspected graphically using histograms Garrido et al. (2014).

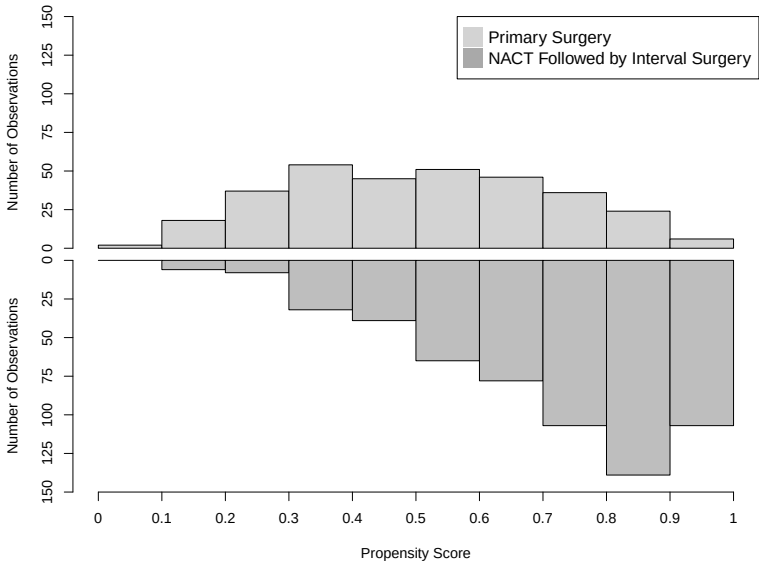


Figure 8.5: Histogram of propensity scores per treatment group

Overall survival by treatment per biomarker subgroup without imputation or weighting

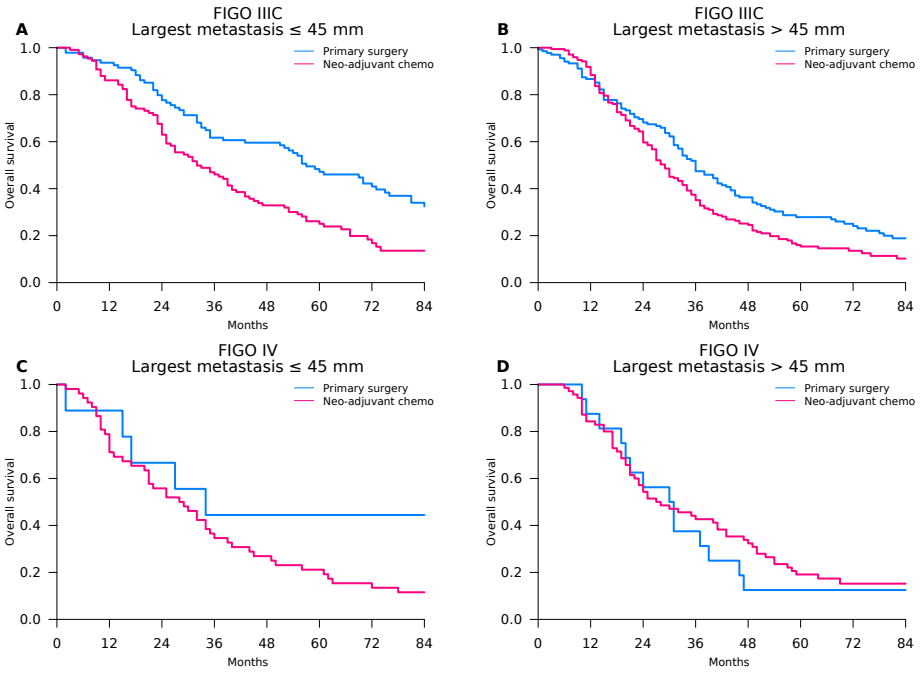


Figure 8.6: Overall survival by treatment group in the biomarker subgroups without imputation or weighting, using only patients with known metastatic tumour size
(A) Subgroup 1 of patients with FIGO stage IIIC and metastatic size ≤ 45mm, HR 1.87, 95% CI (1.35–2.59) $p = 0.0002$;
(B) Subgroup 2 of patients with FIGO stage IIIC and metastatic size > 45mm, HR 1.37, 95% CI (1.07–1.75) $p = 0.013$;
(C) Subgroup 3 of patients with FIGO stage IV and metastatic size ≤ 45mm, HR 1.75, 95% CI (0.74–4.10) $p = 0.20$;
(D) Subgroup 4 of patients with FIGO stage IV and metastatic size > 45mm, HR 0.90, 95% CI (0.50–1.62) $p = 0.72$
The heterogeneity of the treatment effect across the four subgroups was statistically significant (p-value for interaction 0.025).

CHAPTER 9

Future Perspectives

Risk prediction and treatment effect estimation have a long history, and even today new methods are being developed for these purposes. In this thesis we have explored only a small selection of the novel methods that have come available to analyze clinical studies, registry data, and to do combined analyses of these two types of sources of data. Below we discuss a few possible directions for future research.

Mixed sources

It would be worthwhile to further investigate methods that can combine data from multiple sources. We have demonstrated in chapter 7 how a non-randomized study can be combined with a control group from a registry, but it would also be interesting to investigate methods that predict individual risks using data from combinations of randomized studies. Perhaps inspiration can be drawn from methods for network meta-analysis, which are currently available to estimate relative effects between pairs of interventions from networks of studies.

For example, the model presented in chapter 2 predicts the individual treatment effect using data from the Boost/No Boost trial, which compared a 16 Gy radiotherapy boost dose to not giving a boost to patients with early-stage breast cancer. Additionally, in the Young Boost trial a 26 Gy boost (i.e., a higher boost dose) is compared with a 16 Gy boost in the subgroup of younger patients. Even if the boost would not be given today, at least not using the techniques applied in the trial, it could be of interest to be able to combine the two trials, to obtain reliable individualized predictions of the effect of a 26 Gy boost compared with no boost at all.

Registry data

Another opportunity for future research lies in the use of registry data during clinical studies. When a clinical study is being designed, many assumptions need to be made that are often subject to some level of uncertainty. For example, both in single-group and randomized studies, the sample size calculation is based on an expectation of patients' prognosis with standard treatment. Such assumptions could be checked during the study at interim analyses. If the assumptions appear to be unrealistic, that could be a reason to modify the design of the study.

The review of the assumptions and the analysis of the operating characteristics of the study may benefit from registry data. This is especially the case for registries that are updated on a regular basis, because they may be able to provide data that were not available at the start of the study. If there are patients who belong to the target population, are treated concurrently with the study, but do not participate in it, their data can be used to revise the design assumptions for short-term study outcomes.

For longer-term outcomes, it may not be possible to use registry data from newly registered patients who started treatment while the study is ongoing. If follow-up data are regularly updated in the registry, updated data from patients that have previously been treated in the real world before the start of the study, may still prove useful.

Effect estimates

In theory, the estimated treatment effect in a randomized study is equal to the average treatment effect (ATE) in the population. Some of the methods explored in this thesis were especially developed to estimate the ATE from observational data. In practice however, the question is whether the patients who participate in a randomized study are fully representative of the target population: the patients who potentially would get the new treatment in the future. If this is not the case, then the estimate obtained from a randomized study is not equal to the ATE. Is it then fair, or even possible, to demand from statistical methods that incorporate real-world data to provide unbiased estimates of the ATE? It would be useful to do more research on why potentially eligible patients are not invited or decline the invitation to participate. If these reasons could be collected into existing registries, it would be easier to investigate the bias that results from incorporation of real-world data into the analysis of clinical studies. It would also help to investigate how results from studies can be extrapolated better to real-world populations.

Communication

Whatever measure is estimated, and whatever method is used to do that, the results should always be presented in a way that makes interpretation as easy as possible. But on the other hand, any prediction comes with some degree of uncertainty. It is difficult enough to communicate when it comes to numbers and probabilities, let alone to communicate about the uncertainty, especially in case of individualized predictions. Risk predictions are commonly used by doctors in conversations with patients as follows: Among 100 patients with the same characteristics as you have, the model predicts that (for example) 70 of them will be event free at 10 years follow-up.

Some controversy exists as to whether such a prediction (70%, or the number 70 in the example above) should be accompanied with an expression of uncertainty. One point of view is that any presented number that is based on a sample should be accompanied with a confidence interval.

The usual 95% confidence interval, however, is too difficult even for some researchers to explain. Bayesian confidence intervals (also called credible intervals) may be easier to explain. A Bayesian confidence interval could for example indicate that there is a 95% probability that between 61 and 79 patients will be event free. The reality is that the interval of 61–79 holds even if the true event-free probability of 70% were perfectly known without any uncertainty. It is a prediction interval and it can be readily calculated from the binomial distribution.

If the true probability is known to be 70%, then the degree of uncertainty about the number of people that will be event free in the future is determined completely by the fact that the doctor chose the number 100 as most convenient in the conversation with the patient and to explain the meaning of the probability. If the doctor would have presented the known true probability of 70% as "700 out of 1000 patients", then the corresponding interval would have been 671–728.

Perhaps the best the doctor can say is something like this: If I had 100 people identical to you, I would expect that about 70 of them will be event-free 10 years after treatment. And I have no idea whether you will be one of these 70 (Kattan, 2011).

Conclusion

In this thesis, we have explored several statistical methods to analyze treatment strategies. We demonstrated their usefulness in a selection of applications in oncology, utilizing various data sources. These methods are widely available, but the challenges lay in the validity of the assumptions they make, and communication of their results.

Summary

In this thesis statistical methods are explored to analyze treatment strategies from both randomized and non-randomized studies in oncology. The choice of this topic is motivated in **chapter 1**, where a number of reasons are described why such methods have gained importance recently.

In **chapter 2** a model was developed to predict the probability of a local recurrence at 10 years for early-stage breast cancer patients treated with breast-conserving therapy. The model was based on data from the EORTC 22881-10882 Boost/No Boost randomized trial, which compared an additional 'boost' dose to the original tumor bed with standard whole-breast irradiation. A variable-selection procedure was used to reduce the effect of over-fitting and make it easier to use the model in practice. The procedure we developed, kept variables in the model if their absolute effect on the prediction was at least 4 percentage points. The model was presented in the form of a nomogram and an interactive web page. These facilitate straightforward calculation of the predicted effect of the boost dose for individual patients, so that the model may assist in shared decision making.

In **chapter 3** an additional analysis of the Boost/No Boost trial was performed after longer follow-up date had been collected. It was shown that the relative association of the grade of the invasive tumor with recurrence decreased over time. Therefore, patients with high-grade invasive tumors should be monitored closely especially in the first years of follow-up. On the other hand, the relative effect of the presence of DCIS adjacent to the invasive tumor remained nearly constant over time. This finding illustrates the importance of long-term follow-up in clinical trials to estimate absolute effects accurately.

The model in chapter 2 accounted for the non-linear association of age and the risk of recurrence. Because the risk of recurrence was higher for younger patients, the biggest absolute treatment effects were observed in the subgroup of younger patients. This result had been found in an earlier analysis of the Boost/No Boost trial and was the reason to initiate the Young Boost randomized trial. This trial was designed to investigate the efficacy of a boost at a higher dose. One of the expected

side effects of the higher dose is a worse cosmetic outcome. In **chapter 4** the agreement was investigated between cosmetic outcome as interpreted by the patient, the treating physician, and a computer program that analyzed digital photographs of the breasts. The agreement between them was low, but associations were found between cosmetic outcome as reported by the patient and specific cosmetic features used by the software. Also, better patient-reported cosmetic outcome was associated with better quality of life.

In **chapter 5** the side effects in the Young boost trial were compared by randomized treatment group. As expected, the higher boost dose increased the risk of fibrosis and a bad cosmetic outcome. Also some other factors were identified that were associated with worse cosmesis, even after adjustment for the effect of a higher boost. These results may offer opportunities for the development of radiotherapy-treatment strategies so that the cosmetic outcome can be improved in the future.

Developments over time in clinical practice and the outcome of early-stage breast cancer patients are investigated in **chapter 6**. A non-experimental cohort was used that consisted of patients who were treated with breast-conserving therapy at the Netherlands Cancer Institute between 1980 and 2008. Over these years, the use and the dose of the radiotherapy boost declined over time in this cohort, but a small improvement of locoregional control over time was observed. However, this trend over time was not apparent after correction for tumor characteristics, that were unfavorable in earlier time periods.

In **chapter 7** an analysis is presented of the long-term outcome of the single-arm phase-II HOVON-130 trial. In this trial, lenalidomide added to the standard R-CHOP chemotherapy as first-line treatment for patients with newly diagnosed Diffuse Large B-Cell Lymphoma (DLBCL) with a rearrangement of the *MYC* gene (*MYC*-R). The analysis was extended by the addition of a cohort of patients who received the standard R-CHOP treatment. Three different statistical methods were used that reduce treatment-selection bias: multi-variable proportional hazards regression adjusting for baseline differences; one-to-one matching based on the international prognostic index (IPI) score; and inverse probability of treatment weighting (ITPW) using a propensity score for being included in the HOVON-130 trial. The results of the three methods were consistently in favor of the lenalidomide add-on. This supports the use of lenalidomide added to R-CHOP as an additional first-line treatment option for *MYC*-R DLBCL.

In **chapter 8** an external validation is presented of a treatment-selection rule. The rule had been developed to aid in the choice between primary surgery and neoadjuvant chemotherapy for treatment of patients with advanced epithelial ovarian cancer. It was was validated in a non-experimental cohort of patients with advanced-stage epithelial ovarian cancer at the Centre of Gynaecologic Oncology Amsterdam between 1998 and 2012. Possible treatment-selection bias was accounted for by in-

verse probability of treatment weighting (ITPW) using a propensity score for the actual treatment received. To demonstrate that balance was achieved after weighting, we extended 'the standardized difference', which is often used to demonstrate balance after matching, so that it can be used after weighting as well. The extension consisted of incorporating the case weights into the formula, and taking into account that the sizes of the treatment groups were unequal. The results of the validation provided further support for the recommendations made on the basis of the model, and may promote the use of it for making recommendations about treatment options for women with advanced-stage ovarian cancer.

The various analyses presented in this thesis are put into perspective in **chapter 9**, where some ideas are put forward about estimating treatment effects from registry data and mixed data sources, and how to communicate the results. In addition, some possible directions for future research are suggested.

Nederlandse Samenvatting

Dit proefschrift gaat over statistische methodes om behandelstrategieën te analyseren voor zowel gerandomiseerd als niet-gerandomiseerd oncologisch onderzoek. In **hoofdstuk 1** wordt de keuze voor dit onderwerp gemotiveerd en worden diverse redenen beschreven waarom het belang ervan recentelijk is toegenomen.

In **hoofdstuk 2** wordt een model ontwikkeld dat de kans voorspelt op een lokaal recidief voor patiënten die vroegstadium borstkanker hebben en behandeld worden met borst-sparende therapie. Voor het model zijn gegevens gebruikt uit de gerandomiseerde EORTC 22881-10882 Boost/No Boost studie, waarin de toevoeging van een radiotherapie 'boost' dosis werd vergeleken met de standaard bestraling van de hele borst. Om het probleem van over-fitting tegen te gaan, en om het model gemakkelijker te maken in het gebruik, is een selectie gemaakt van de variabelen die in het model zijn opgenomen. De procedure die we hebben ontworpen, behoudt alleen die variabelen in het model waarvan het absolute effect op de voorspelde kans tenminste 4 procentpunt is. Het model is gepresenteerd in de vorm van een nomogram (dat wil zeggen een lijnfiguur) en een interactieve webpagina. Daarmee kan eenvoudig worden berekend wat het voorspelde behandeleffect van de boost dosis is voor een bepaalde patiënt, zodat het model kan worden gebruikt bij de gezamenlijke besluitvorming van arts en patiënt.

In **hoofdstuk 3** wordt een vervolganalyse gepresenteerd van de Boost/No Boost studie, die gebruik maakt van gegevens die beschikbaar zijn gekomen nadat de patiënten uit die studie langer waren gevolgd. Uit deze analyse blijkt dat de samenhang tussen de differentiatiegraad van de invasieve borsttumor en het risico op recidief afneemt gedurende het verloop van tijd. Daarom wordt aanbevolen om patiënten die een tumor hebben met een hoge graad intensief te volgen gedurende de eerste tijd na de behandeling. Daar staat tegenover dat het relatieve effect van de aanwezigheid van DCIS naast de invasieve tumor vrijwel gelijk blijft in de tijd. Deze bevinding illustreert dat het belangrijk is om deelnemers van medisch wetenschappelijke onderzoeken voldoende lang te volgen om absolute effecten van behandeling goed te kunnen schatten.

In het nomogram uit hoofdstuk 2 was de samenhang tussen leeftijd en risico op recidief gemodelleerd volgens een niet-lineair verband, waarbij voor jongere patiënten het risico op recidief groter is dan voor oudere. Daardoor is het behandeleffect van de boost het grootst voor jongere patiënten. Dit resultaat was al gevonden in een eerdere analyse van de Boost/No Boost studie, en was toen aanleiding om een vervolgstudie op te zetten. In deze vervolgstudie, genaamd de Young Boost studie, wordt de effectiviteit onderzocht van extra hoge boost dosis. Een van de verwachte bijwerkingen van een verdere dosisverhoging is een slechtere cosmetische uitkomst. In **hoofdstuk 4** wordt onderzocht in hoeverre de score van de cosmetische uitkomst door de patiënt zelf, overeenkomt met de score die wordt gerapporteerd door de arts, en de score berekend door een computerprogramma dat digitale foto's van de borsten analyseert. De voor kans gecorrigeerde mate van overeenkomst was laag, maar er werd wel gevonden dat de door de patiënt gerapporteerde cosmetiek samenhangt met specifieke kenmerken die door de software worden berekend. Ook kon er een verband worden aangetoond tussen kwaliteit van leven en de cosmetische uitkomst zoals ervaren door de patiënt.

In **hoofdstuk 5** worden de bijwerkingen vergeleken van de twee behandelingen in de Young Boost studie. Zoals verwacht, verhoogt de hoge boost dosis het risico op fibrose en een slechte cosmetische uitkomst vergeleken met de lage boost dosis. Daarnaast kunnen er andere factoren worden geïdentificeerd die samenhangen met slechte cosmetiek, zelfs na correctie voor het effect van de hogere boost dosis. Deze kunnen mogelijk in de toekomst worden gebruikt voor de ontwikkeling van nieuwe bestralingsmethoden met als doel de cosmetische uitkomst te verbeteren.

In **hoofdstuk 6** worden ontwikkelingen in de tijd onderzocht in de behandeling en de behandeluitkomsten voor patiënten met vroegstadium borstkanker. Hiervoor is een niet-experimenteel cohort gebruikt van patiënten die behandeld zijn met borstsparende therapie in het Antoni van Leeuwenhoek ziekenhuis in de periode van 1980 tot 2008. In de loop van deze periode daalde het relatieve aantal patiënten dat met een boost dosis werd bestraald, en als een boost werd gegeven werd daarvoor steeds minder vaak een hoge dosis voor gebruikt. Niettemin daalde de locoregionale controle niet gedurende deze periode, maar verbeterde zelfs—zij het in geringe mate. Deze verbetering was echter niet aanwijsbaar na correctie voor kenmerken van de ziekte, die in het begin van de onderzochte periode minder gunstig waren.

In **hoofdstuk 7** wordt een analyse gepresenteerd van de lange-termijnuitkomsten van de HOVON-130 studie. Deze studie is een fase II onderzoek zonder controlegroep, waarin patiënten met diffuus grootcellige B-cel lymfomen (DLBCL) die een breuk hebben in het *MYC* oncogen, lenalidomide kregen toegevoegd aan behandeling met R-CHOP. De analyse van deze studie is uitgebreid door er gegevens aan toe te voegen van een cohort dat bestaat uit patiënten die de standaard R-

CHOP behandeling hebben gekregen. Schattingen van behandelingseffecten waarbij niet volledig vergelijkbare patiëntengroepen worden gebruikt, zijn statistisch niet zuiver. In de analyse is getracht deze onzuiverheid ten gevolge van zulke behandelingsselectie te verminderen. Daartoe is voor drie statistische methoden onderzocht, hoe de keuze van methode de resultaten van de analyse beïnvloedt: correctie voor verschil in kenmeren van de patiënten bij aanvang van behandeling door middel van multipele regressie in het evenredig-risicomodel van Cox; constructie van een verzameling van gekoppelde patiënten ('matching'), waarbij paren worden gevormd van patiënten die verschillend behandeld zijn, maar dezelfde score hebben volgens de internationale prognostische index (IPI); en tot slot het wegen van patiënten met het omgekeerde van de kans op behandeling ('inverse probability of treatment weighting', IPTW) door middel van een 'propensity score' voor deelname aan de HOVON-130 studie. De overeenstemming van de resultaten van de drie methoden, ten gunste van toevoegen van lenalidomide, ondersteunt de conclusie dat voor deze patiënten lenalidomide toegevoegd aan R-CHOP een goede behandelmogelijkheid is in de eerste lijn.

In **hoofdstuk 8** wordt de externe validiteit onderzocht van een behandelkeuzestrategie. Deze strategie was eerder ontwikkeld en kan worden gebruikt om uit patiënten met vergevorderde epitheliale eierstokkanker, te selecteren wie het meeste baat heeft bij neoadjuvante chemotherapie, en wie beter meteen kan worden geopereerd ('primaire chirurgie'). De strategie wordt gevalideerd in een niet-experimenteel cohort van patiënten die behandeld zijn in het Centrum voor Gynaecologische Oncologie Amsterdam in de periode van 1998 tot 2012. De mogelijke onzuiverheid ten gevolge van de manier waarop de behandelkeuze tot stand kwam, is verminderd door middel van IPTW met de propensity score. Om aannemelijk te maken dat de onzuiverheid is verminderd, is gebruik gemaakt van gestandaardiseerde verschillen. Dit concept, dat vaak wordt gebruikt in analyse van gematchte paren, is voor dit doel uitgebreid zodat het ook toegepast kan worden voor gewogen analyses. Dit is gedaan door in de formule van het gestandaardiseerde verschil de gewichten op te nemen, en er rekening mee te houden dat de behandelgroepen niet even groot hoeven te zijn. De resultaten van deze validatie onderschrijven de aanbevelingen volgens de strategie, en kunnen het gebruik ervan bevorderen bij de keuze tussen de behandelopties voor patiënten met vergevorderde eierstokkanker.

De diverse analyses die gepresenteerd worden in dit proefschrift worden in perspectief geplaatst in **hoofdstuk 9**. Daarin worden enkele ideeën geopperd over het schatten van behandelingseffecten met behulp van registers en combinaties van gegevensbronnen, en communicatie van resultaten. Daarnaast worden enkele suggesties gedaan voor mogelijke richtingen van toekomstig onderzoek.

PhD Portfolio

	Institute	Year
General courses		
The AMC World of Science	AMC Graduate School	2015
Bioinformatics	AMC Graduate School	2016
English Writing and Presenting in Biomedicine	Oncology Graduate School Amsterdam	2016
BROK: AvL specifiek	NKI-AvL	2018
Medical Device Investigations vs. Pharma Trials	NKI-AvL	2020
Oncologisch Spectrum Hematologie	IKNL	2021
Basic ICH-GCP R2 training EMA/CHMP/ICH/135/1995	HOVON	2021
Specific courses		
Bayesian Modeling for Cognitive Science	University of Amsterdam	2011
Survival and Event History Analysis	University of Milano – Bicocca	2014
Methods for Clinical and Translational Research	Johns Hopkins Bloomberg School of Public Health	2017
Bayesian Adaptive Trials	Johns Hopkins Bloomberg School of Public Health	2017
Biostatistical Analysis of Epidemiologic Data II: Poisson and Conditional Logistic Regression Analysis	Johns Hopkins Bloomberg School of Public Health	2017
Multilevel Models	Johns Hopkins Bloomberg School of Public Health	2017
Introduction to Causal Inference	University of Florence	2021
Statistical & Practical Aspects of the Design and Analysis of Multi Arm Multi-Stage Platform Trials	UCL Institute of Clinical Trials and Methodology	2021

	Institute	Year
Seminars, workshops, and master classes		
Special Topics on Sequential Methodology	UMC Utrecht	2013
NVMETC Scholingsdag	NVMETC	2015
Small-Population Research Methods Projects and Regulatory Application Workshop	EMA	2017
Early Breast Cancer Trialists' Collaborative Group (EBCTCG)	Oxford University	2018
Randomization and Stratification	Vrouw Kind Centrum (AMC)	2019
The Future of Clinical Trials and Evidence Generation, and Their Use in Regulatory Decision Making	Regulatory Science Network Netherlands (RSNN)	2019
Data and Safety Monitoring Boards	DORP	2021
Presentations		
Preliminary Results of the Young Boost Trial and Validation of the IBTR Nomogram for BCT	Aarhus University	2013
Treatment-Selection Markers: Benefit and Indication Bias	MEMTAB Conference	2016
Development of Treatment-Selection Markers for Studies with Failure-Time Outcomes	ISCB	2017
Development of a Treatment-Selection Rule for Choosing Between Two Chemotherapy Treatments From a Randomized Trial	ISCB	2019
Awards		
Louise Gunning Public Health Study Fund	AMC Young Talent Fund	2017

Publications

Part of this thesis

Erik van Werkhoven, Guus Hart, Harm van Tinteren, Paula Elkhuizen, Laurence Collette, Philip Poortmans, and Harry Bartelink. Nomogram to predict ipsilateral breast relapse based on pathology review from the EORTC 22881-10882 boost versus no boost trial. *Radiotherapy and Oncology*, 100(1):101 – 107, 2011.

Conny Vrieling[*], **Erik van Werkhoven**[*], Philippe Maingon, Philip Poortmans, Caroline Weltens, Alain Fourquet, Dominic Schinagl, Bing Oei, Carla C. Rodenhuis, Jean-Claude Horiot, Henk Struikmans, Erik van Limbergen, Youlia Kirova, Paula Elkhuizen, Rudolf Bongartz, Raymond Miralbell, David A. L. Morgan, Jean-Bernard Dubois, Vincent Remouchamps, René-Olivier Mirimanoff, Guus Hart, Sandra Collette, Laurence Collette, and Harry Bartelink. Prognostic factors for local control in breast cancer after long-term follow-up in the EORTC boost vs no boost trial: A randomized clinical trial. *JAMA Oncology*, 3(1):42–48, 2017.

Patricia J. A. M. Brouwers[*], **Erik van Werkhoven**[*], Harry Bartelink, Alain Fourquet, Claire Lemanski, Judith van Loon, John H. Maduro, Nicola S. Russell, Luc J. E. E. Scheijmans, Dominic A. X. Schinagl, Antonia H. Westenberg, Philip Poortmans[†], and Liesbeth J. Boersma[†] on behalf of the Young Boost Trial research group. Factors associated with patient-reported cosmetic outcome in the young boost breast trial. *Radiotherapy and Oncology*, 120(1):107–113, 2016.

Patricia J. A. M. Brouwers[*], **Erik van Werkhoven**[*], Harry Bartelink, Alain Fourquet, Claire Lemanski, Judith van Loon, John H. Maduro, Nicola S. Russell, Luc J. E. E. Scheijmans, Dominic A. X. Schinagl, Antonia H. Westenberg, Philip Poortmans[†], Liesbeth J. Boersma[†], and Young Boost Trial research group. Predictors for poor cosmetic outcome in patients with early stage breast cancer treated with breast conserving therapy: Results of the young boost trial. *Radiotherapy and Oncology*, 128(3):434–411 2018.

S. C. J. Bosma[*], F. van der Leij[*], **E. van Werkhoven**[*], H. Bartelink, J. Wesseling, S. Linn, E. J. Rutgers, M. J. van de Vijver, and P. H. M. Elkhuizen. Very low local recurrence rates after breast-conserving therapy: analysis of 8485 patients treated over a 28-year period. *Breast Cancer Research and Treatment*, 156(2):391–400, 2016.

R. van de Vrie[*], **E. van Werkhoven**[*], J. D. Asseler M. J. Rutten, H. S. van Meurs, D. E. Werter, A. H. Zwinderman M. R. Buist, C. A. R. Lok, P. M. M. Bossuyt, G. G. Kenter, and P. Tajik. Validation of a treatment-selection rule for patients with advanced stage ovarian cancer. *European Journal of Gynaecological Oncology*, 41(6):1023–1030, 2020.

*) Equally contributing

Selected Other Publications

Femke van der Leij, **Erik van Werkhoven**, Sophie Bosma, Sabine C. Linn, Emiel J. Rutgers, Marc J. van de Vijver, Harry Bartelink, Paula H. M. Elkhuizen†, and Astrid Scholten†. Low risk of recurrence in elderly patients treated with breast conserving therapy in a single institute. *The Breast*, 30:19–25, 2016.

Erik van Werkhoven, Parvin Tajik, and Patrick M. Bossuyt. Always randomize as late as possible. *Gastric Cancer*, 6:1308–1309, 2019.

Erik van Werkhoven, Samantha Hinsley, Eleni Frangou, Jane Holmes, Rosemarie de Haan, Maria Hawkins, Sarah Brown, and Sharon B Love. Practicalities in running early-phase trials using the time-to-event continual reassessment method (TiTE-CRM) for interventions with long toxicity periods using two radiotherapy oncology trials as examples. *BMC Medical Research Methodology*, 20, 2020.

Marieke E. M. van der Noordaa, Ileana Ioan, Emiel J. Rutgers, **Erik van Werkhoven**, Claudette E. Loo, Rosie Voorthuis, Jelle Wesseling, Japke van Urk, Terry Wiersma, Vincent Dezentje, Marie-Jeanne T. F. D. Vrancken Peeters, and Frederieke H. van Duijnhoven. Breast-conserving therapy in patients with cT3 breast cancer with good response to neoadjuvant systemic therapy results in excellent local control: A comprehensive cancer center experience. *Annals of Surgical Oncology*, 28:7383–7394, 2021.

Sophie C. J. Bosma, Marlous Hoogstraat, **Erik van Werkhoven**, Michiel de Maaker, Femke van der Leij, Paula H. M. Elkhuizen, Alain Fourquet, Philip Poortmans, Liesbeth J. Boersma, Harry Bartelink, and Marc J. van de Vijver. A case-control study to identify molecular risk factors for local recurrence in young breast cancer patients. *Radiotherapy and Oncology*, 156:127–135, 2021.

Dankwoord

Dit proefschrift had niet tot stand kunnen komen zonder de artsen, onderzoekers en anderen met wie ik heb mogen samenwerken en van wie ik veel heb kunnen leren, en natuurlijk ook niet zonder de vele patiënten die toestemming gegeven hebben voor het gebruik van hun gegevens voor de onderzoeken die in dit proefschrift besproken worden.

Mijn promotoren, prof. Bossuyt en prof. Zwinderman, hartelijk dank voor alles wat jullie hebben gedaan, en de mogelijkheden die jullie me hebben gegeven. Jullie wijsheid, vriendelijkheid en geduld zijn een voorbeeld voor mij. Koos, ik denk met plezier terug aan de gesprekken die we hadden, o.a. bij het schoolbord op je kamer. Patrick, ik bewonder de manier waarop je ieder gesprek snel tot de kern kunt brengen, ongeacht het onderwerp lijkt het wel. Dank je wel voor de onvermoeibaarheid waarmee je me bleef aanmoedigen.

Dank aan de commissieleden, prof. Besselink, prof. Le Cessie, prof. Cornelissen, dr. Van Oijen, prof. Roes en prof. Stalpers voor uw bereidheid om zitting te nemen in de promotiecommissie en dit proefschrift kritisch te beoordelen.

Prof. Bartelink, Harry, ik bewonder de manier waarop het je lukt om groepen eigenwijze wetenschappers bij elkaar te brengen en consesus te creëren. Enorm bedankt voor je steun en vertrouwen.

Guus Hart, dank je wel voor je hulp en je creativiteit bij het benaderen van problemen, o.a. bij de boost/no-boost studie. Harm van Tinteren, dank voor de vrijheid die je me gaf op de afdeling Biometrics van het Antoni van Leeuwenhoek. Het is fijn dat we in de commissie BOM nog samen kunnen sparren.

Prof. Poortmans, Philip, en prof. Boersma, Liesbeth, dank jullie wel voor jullie steun en prettige manier van samenwerken. Astrid Scholten, heel fijn dat je vanuit het AvL betrokken wilt zijn bij de Young Boost studie. Sophie Bosma, dank je wel voor je inzet voor de Young boost trial en de samenwerking bij het tumor-registratie cohort, samen met Paula Elkhuizen en Femke van der Leij. Dank ook aan iedereen van de tumorregistratie en de afdeling Biometrics van het NKI-AvL.

Prof. Linn, Sabine, prof. Wesseling, Jelle, en prof. Van de Vijver, Marc, dank voor jullie bijdrage niet alleen aan dit project, maar ook voor jullie hulp om mijn proefschrift mogelijk te maken. Verder dank aan iedereen die ik nog niet genoemd heb van de EORTC Radiation Oncology and Breast Cancer Groups en Young Boost Trial research group, waaronder Conny Vrieling, Laurence Collette, Sandra Collette, Patricia Brouwers, John Maduro, en Nicola Russell.

Vera de Jonge en Martine Chamuleau, enorm bedankt voor de prettige samenwerking, en ook Marcel Nijland, Avinash Dinmohamed, prof. Kersten, Marie José en prof. De Jong, Daphne, en alle andere mede auteurs van de publicatie van de HOVON-130 studie, dank voor jullie waardevolle bijdrages.

Parvin Tajik, Roelien van de Vrie, en alle andere mede auteurs waaronder Christianne Lok en prof. Kenter, Gemma, dank voor jullie vertrouwen en de prettige samenwerking.

Verder wil ik het Louise Gunning Public Health Study Fund bedanken voor het mogelijk maken van mijn bezoek aan de Johns Hopkins Bloomberg School of Public Health in Baltimore en het Fred Hutchinson Cancer Center in Seattle. Holly Janes, thank you so much for your hospitality!

Dank ook aan mijn collega's Karolina Sikorska, Marta López Yurda, Rob Kessels, Vincent van der Noort, en alle anderen van de afdeling Biometrics van het NKI-AvL. Daarnaast mijn collega's in Amsterdam, waar ik me altijd welkom heb gevoeld: Parvin Tajik, Mariska Leeflang, Jungfeng Wang, en alle anderen van de BiTE groep. En ook dank aan mijn nieuwe collega's bij de HOVON: Dana Chițu, Esther Oomen-de Hoop, Kazem Nasserinejad, Patrycja Gradowska, Ronnie van der Holt, Yvette van Norden, Sjoerd Hermans, Niek van der Maas, en Jurjen Versluis. Prof. Cornelissen, beste Jan, veel dank voor de mogelijkheid die je me hebt geboden om dit proefschrift af te maken.

Tenslotte wil ik graag al mijn vrienden bedanken en mijn familie. In het bijzonder mijn paranimfen Daniël Worm en Vincent van der Noort. Maar bovenal mijn ouders, die me mijn leven lang overal in hebben gesteund.

Bibliography

Aaronson, N. K., Ahmedzai, S., Bergman, B., et al. 1993. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *J. Natl. Cancer Inst.*, 85:365–376 (p. 46).

Abramson, J. S., Ruppert, A. S., Giri, S., et al. 2021. randomized phase II/III study of DA-EPOCH-R +/- venetoclax in previously untreated double hit lymphoma: Initial results from Alliance A051701. *Blood*, 138:523–523 (p. 95).

Al Uwini, S., Antonini, N., Poortmans, P. M., et al. 2009. The influence of the use of CT-planning on the irradiated boost volume in breast conserving treatment. *Radiother. Oncol.*, 93:87–93 (p. 69).

Anderson, S. J., Wapnir, I., Dignam, J. J., et al. 2009. prognosis after ipsilateral breast tumor recurrence and locoregional recurrences in patients treated by breast-conserving therapy in five national surgical adjuvant breast and bowel project protocols of node-negative breast cancer. *J. Clin. Oncol.*, 27:2466–2473 (p. 91).

Antonini, N., Jones, H., Horiot, J. C., et al. 2007. Effect of age and radiation dose on local control after breast conserving treatment: EORTC trial 22881-10882. *Radiother. Oncol.*, 82:265–271 (pp. 7, 13).

Augustine, E. F., Adams, H. R., & Mink, J. W. 2013. Clinical trials in rare disease: Challenges and opportunities. *J. Child Neurol.*, 28:1142–1150 (p. 2).

Aukema, S. M., Kreuz, M., Kohler, C. W., et al. 2014. Biological characterization of adult MYC-translocation-positive mature B-cell lymphomas other than molecular Burkitt lymphoma. *Haematologica*, 99:726–735 (p. 95).

Austin, P. C. 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat. Med.*, 28:3083–3107 (p. 129).

—. 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar. Behav. Res.*, 46:399–424 (p. 97).

Austin, P. C., & Stuart, E. A. 2015. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat. Med.*, 34:3661–3679 (p. 108).

Bantema-Joppe, E. J., Schilstra, C., de Bock, G. H., et al. 2012. Simultaneous integrated boost irradiation after breast-conserving surgery: Physician-rated toxicity and cosmetic outcome at 30 months' follow-up. *Int. J. Radiat. Oncol. Biol. Phys.*, 83:e471–e477 (p. 71).

Barnett, G. C., Wilkinson, J. S., Moody, A. M., et al. 2011. The cambridge breast intensity-modulated radiotherapy trial: Patient- and treatment-related factors that influence late toxicity. *Clin. Oncol.*, 23:662–673 (pp. 59, 70, 71).

Bartelink, H., Bourgier, C., & Elkhuizen, P. 2012. Has partial breast irradiation by IORT or brachytherapy been prematurely introduced into the clinic? *Radiother. Oncol.*, 104:139–142 (pp. 32, 90).

Bartelink, H., Horiot, J.-C., Poortmans, P., et al. 2001. Recurrence rates after treatment of breast cancer with standard radiotherapy with or without additional radiation. *New Engl. J. Med.*, 345:1378–1387 (pp. 7, 25).

Bartelink, H., Horiot, J.-C., Poortmans, P. M., et al. 2007. Impact of a higher radiation dose on local control and survival in breast-conserving therapy of early breast cancer: 10-year results of the randomized boost versus no boost EORTC 22881-10882 trial. *J. Clin. Oncol.*, 25:3259–3265 (pp. 7, 8, 25, 59, 92).

Bartelink, H., Maingon, P., Poortmans, P., et al. 2015. Whole-breast irradiation with or without a boost for patients treated with breast-conserving surgery for early breast cancer: 20-year follow-up of a randomised phase 3 trial. *Lancet Oncol.*, 16:47–56 (pp. 24, 25, 31, 43, 90).

Beadle, B. M., Woodward, W. A., & Buchholz, T. A. 2011. The impact of age on outcome in early-stage breast cancer. *Semin. Radiat. Oncol.*, 21:26–34 (pp. 7, 13).

Bellon, J. R. 2015. Personalized radiation oncology for breast cancer: The new frontier. *J. Clin. Oncol.*, 33:1998–2000 (p. 24).

Bentzon, N., Düring, M., Rasmussen, B. B., Mouridsen, H., & Kroman, N. 2008. Prognostic effect of estrogen receptor status across age in primary breast cancer. *Int. J. Cancer*, 122:1089–1094 (p. 92).

Blamey, R. W., Bates, T., Chetty, U., et al. 2013. Radiotherapy or tamoxifen after conserving surgery for breast cancers of excellent prognosis: British Association of Surgical Oncology (BASO) II trial. *Eur. J. Cancer*, 49:2294–2302 (p. 92).

Boersma, L. J., Janssen, T., Elkhuizen, P. H. M., et al. 2012. Reducing interobserver variation of boost-CTV delineation in breast conserving radiation therapy using a pre-operative CT and delineation guidelines. *Radiother. Oncol.*, 103:178–182 (pp. 72, 73).

Bollet, M. A., Sigal-Zafrani, B., Mazeau, V., et al. 2007. Age remains the first prognostic factor for loco-regional breast cancer recurrence in young (<40 years) women treated with breast conserving surgery first. *Radiother. Oncol.*, 82:272–280 (pp. 7, 13, 90, 92).

Bontenbal, M., Nortier, J. W., Beex, L. V., et al. 2000. Adjuvant systemic therapy for patients with resectable breast cancer: Guideline from the Dutch National Breast Cancer Platform and the Dutch Society for Medical Oncology. *Ned. Tijdschr. Geneeskd.*, 144:984–989 (p. 81).

Borger, J., Kemperman, H., Hart, A., et al. 1994. Risk factors in breast-conservation therapy. *J. Clin. Oncol.*, 12:653–660 (p. 81).

Bosco, J. L. F., Silliman, R. A., Thwin, S. S., et al. 2010. A most stubborn bias: No adjustment method fully resolves confounding by indication in observational studies. *J. Clin. Epidemiol.*, 63:64–74 (p. 1).

Bouganim, N., Tsvetkova, E., Clemons, M., & Amir, E. 2013. Evolution of sites of recurrence after early breast cancer over the last 20 years: Implications for patient care and future research. *Breast Cancer Res. Treat.*, 139:603–606 (p. 90).

Bristow, R. E., Eisenhauer, E. L., Santillan, A., & Chi, D. S. 2007. Delaying the primary surgical effort for advanced ovarian cancer: A systematic review of neoadjuvant chemotherapy and interval cytoreduction. *Gynecol. Oncol.*, 104:480–490 (p. 117).

Brouwers, P. J. A. M., van Werkhoven, E., Bartelink, H., et al. 2016. Factors associated with patient-reported cosmetic outcome in the Young Boost breast trial. *Radiother. Oncol.*, 120:107–113 (p. 59).

Brunault, P., Suzanne, I., Trzepidur-Edom, M., et al. 2013. Depression is associated with some patient-perceived cosmetic changes, but not with radiotherapy-induced late toxicity, in long-term breast cancer survivors. *Psycho-oncology*, 22:590–597 (p. 49).

Cabioglu, N., Hunt, K. K., Buchholz, T. A., et al. 2005. Improving local control with breast-conserving therapy: A 27-year single-institution experience. *Cancer*, 104:20–29 (pp. 79, 90).

Cardoso, J. S., & Cardoso, M. J. 2007. Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. *Artif. Intell. Med.*, 40:115–126 (pp. 44, 61).

Cardoso, M. J., Cardoso, J., Amaral, N., et al. 2007. Turning subjective into objective: The BCCT.core software for evaluation of cosmetic results in breast cancer conservative treatment. *Breast*, 16:456–461 (pp. 43, 44, 50, 61).

Chagpar, A. B., Killelea, B. K., Tsangaris, T. N., et al. 2015. A randomized, controlled trial of cavity shave margins in breast cancer. *New Engl. J. Med.*, 373:503–510 (p. 91).

Chamuleau, M., Nijland, M., Lamers, N., et al. 2017. First report on a successful screening program for *MYC* Rearrangements and a prospective clinical trial based on MYC Rearrangement in newly diagnosed DLBCL patients in the Netherlands. *Blood*, 130:4144 (p. 96).

Chamuleau, M. E. D., Burggraaff, C. N., Nijland, M., et al. 2020. Treatment of patients with MYC rearrangement positive large B-cell lymphoma with R-CHOP plus lenalidomide: Results of a multicenter HOVON phase II trial. *Haematologica*, 105:2805–2812 (pp. 96, 107, 108).

Chen, W., Sonke, J.-J., Stroom, J., et al. 2015. The effect of age in breast conserving therapy: A retrospective analysis on pathology and clinical outcome data. *Radiother. Oncol.*, 114:314–321 (p. 24).

Chi, D. S., Bristow, R. E., Armstrong, D. K., & Karlan, B. Y. 2011. Is the easier way ever the better way? *J. Clin. Oncol.*, 29:4073–4075 (p. 117).

Christie, D., O'Brien, M., Christie, J., et al. 1996. A comparison of methods of cosmetic assessment in breast conservation treatment. *Breast*, 5:358–367 (p. 49).

Cohen, J. 1968. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.*, 70:213–220 (p. 46).

Coiffier, B., Lepage, E., Briere, J., et al. 2002. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *New Engl. J. Med.*, 346:235–242 (p. 95).

Collette, S., Collette, L., Budiharto, T., et al. 2008. Predictors of the risk of fibrosis at 10 years after breast conserving therapy for early breast cancer–a study based on the EORTC Trial 22881-10882 'boost *versus* no boost'. *Eur. J. Cancer*, 44:2587–2599 (pp. 7, 13, 19, 32, 59, 70).

Collins, G. S., de Groot, J. A., Dutton, S., et al. 2014. External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC Med. Res. Methodol.*, 14:40 (p. 127).

Committee for Medicinal Products for Human Use (CHMP). 2006, Guideline on clinical trials in small populations (p. 2).

Coulter, A., & Collins, A. 2011, Making shared decision-making a reality: No decision about me, without me, available from `https://www.kingsfund.org.uk/publications/making-shared-decision-making-reality`, Retrieved 28 July 2022 (p. 1).

Van Dongen, J. A., Voogd, A. C., Fentiman, I. S., et al. 2000. Long-term results of a randomized trial comparing breast-conserving therapy with mastectomy: European Organization for Research and Treatment of Cancer 10801 trial. *J. Natl. Cancer Inst.*, 92:1143–1150 (pp. 79, 90).

van Driel, W. J., Koole, S. N., Sikorska, K., et al. 2018. Hyperthermic Intraperitoneal Chemotherapy in Ovarian Cancer. *New Engl. J. Med.*, 378:230–240 (p. 118).

Drukker, C. A., Elias, S. G., Nijenhuis, M. V., et al. 2014. Gene expression profiling to predict the risk of locoregional recurrence in breast cancer: A pooled analysis. *Breast Cancer Res. Treat.*, 148:599–613 (p. 92).

Dunleavy, K., Fanale, M. A., Abramson, J. S., et al. 2018. Dose-adjusted EPOCH-R (etoposide, prednisone, vincristine, cyclophosphamide, doxorubicin, and rituximab) in untreated aggressive diffuse large B-cell lymphoma with *MYC* rearrangement: A prospective, multicentre, single-arm phase 2 study. *Lancet Haematol.*, 5:e609–e617 (p. 95).

Early Breast Cancer Trialists' Collaborative Group (EBCTCG). 2005. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: An overview of the randomised trials. *Lancet*, 366:2087–106 (pp. 7, 59).

—. 2011. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: Meta-analysis of individual patient data for 10,801 women in 17 randomised trials. *Lancet*, 378:1707–1716 (p. 79).

—. 2012. Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet*, 379:432–444 (p. 79).

Elkhuizen, P. H., van Slooten, H. J., Clahsen, P. C., et al. 2000. High local recurrence risk after breast-conserving therapy in node-negative premenopausal breast cancer patients is greatly reduced by one course of perioperative chemotherapy: A European Organization for Research and Treatment of Cancer Breast Cancer Cooperative Group Study. *J. Clin. Oncol.*, 18:1075–1083 (p. 7).

Elkhuizen, P. H. M., van de Vijver, M. J., Hermans, J., et al. 1998. Local recurrence after breast-conserving therapy for invasive breast cancer: High incidence in young patients and association with poor survival. *Int. J. Radiat. Oncol. Biol. Phys.*, 40:859–867 (p. 90).

Elston, C. W., & Ellis, I. O. 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology*, 19:403–410 (p. 25).

Eriksson, L., Czene, K., Rosenberg, L., Humphreys, K., & Hall, P. 2013. Possible influence of mammographic density on local and locoregional recurrence of breast cancer. *Breast Cancer Res.*, 15:R56 (p. 24).

Ernst, M. F., Voogd, A. C., Coebergh, J.-W. W., Poortmans, P. M., & Roukema, J. A. 2004. Using locoregional recurrence as an indicator of the quality of breast cancer treatment. *Eur. J. Cancer*, 40:487–493 (p. 91).

European Commission Regulation. 2006, (EC) 507/2006 on the conditional marketing authorisation for medicinal products for human use falling within the scope of Regulation (EC) No 726/2004 of the European Parliament and of the Council, OJ L 92/6, `https://www.gmp-compliance.org/files/guidemgr/reg_2006_507_en.pdf` (p. 2).

Fagotti, A., Ferrandina, G., Fanfani, F., et al. 2008. Prospective validation of a laparoscopic predictive model for optimal cytoreduction in advanced ovarian carcinoma. *Am. J. Obstet. Gynecol.*, 199:642.e1–642.e6 (p. 127).

Fisher, B., Anderson, S., Bryant, J., et al. 2002. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *New Engl. J. Med.*, 347:1233–1241 (p. 79).

Fitzal, F., Krois, W., Trischler, H., et al. 2007. The use of a breast symmetry index for objective evaluation of breast cosmesis. *Breast*, 16:429–435 (p. 43).

Foster, J. C., Freidlin, B., Kunos, C. A., & Korn, E. L. 2020. Single-Arm Phase II Trials of Combination Therapies: A Review of the CTEP Experience 2008–2017. *J. Natl. Cancer Inst.*, 112:128–135 (p. 2).

Franco, P., Cante, D., Sciacero, P., et al. 2015. Tumor bed boost integration during whole breast radiotherapy: A review of the current evidence. *Breast Care*, 10:44–49 (p. 71).

Furet, E., Peurien, D., Fournier-Bidoz, N., et al. 2014. Plastic surgery for breast conservation therapy: How to define the volume of the tumor bed for the boost? *Eur. J. Surg. Oncol.*, 40:830–834 (p. 72).

Fyles, A. W., McCready, D. R., Manchul, L. A., et al. 2004. Tamoxifen with or without breast irradiation in women 50 years of age or older with early breast cancer. *New Engl. J. Med.*, 351:963–970 (p. 32).

Gaffney, D. K., Tsodikov, A., & Wiggins, C. L. 2003. Diminished survival in patients with inner versus outer quadrant breast cancers. *J. Clin. Oncol.*, 21:467–472 (p. 92).

Garrido, M. M., Kelley, A. S., Paris, J., et al. 2014. Methods for constructing and assessing propensity scores. *Heal. Serv. Res.*, 49:1701–1720 (p. 133).

Gönen, M., & Heller, G. 2005. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*, 92:965–970 (p. 9).

González Sanchis, A., Brualla González, L., Fuster Diana, C., et al. 2013. Tumor bed segmentation: First step for partial breast irradiation. *Clin. & Transl. Oncol.*, 15:39–45 (p. 72).

Gopalakrishnan, R., Matta, H., Tolani, B., Triche, T., & Chaudhary, P. M. 2016. Immunomodulatory drugs target IKZF1-IRF4-MYC axis in primary effusion lymphoma in a cereblon-dependent manner and display synergistic cytotoxicity with BRD4 inhibitors. *Oncogene*, 35:1797–1810 (p. 96).

Guinot, J. L., Roldan, S., Maroñas, M., et al. 2007. Breast-conservative surgery with close or positive margins: Can the breast be preserved with high-dose-rate brachytherapy boost? *Int. J. Radiat. Oncol. Biol. Phys.*, 68:1381–1387 (p. 17).

Habermann, T. M., Weller, E. A., Morrison, V. A., et al. 2006. Rituximab-CHOP versus CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. *J. Clin. Oncol.*, 24:3121–3127 (p. 95).

Haloua, M. H., Krekel, N. M. A., Jacobs, G. J. A., et al. 2014. cosmetic outcome assessment following breast-conserving therapy: A comparison between BCCT.core software and panel evaluation. *Int. J. Breast Cancer*, 2014:716860 (p. 52).

Hammer, C., Maduro, J. H., Bantema-Joppe, E. J., et al. 2017. Radiation-induced fibrosis in the boost area after three-dimensional conformal radiotherapy with a simultaneous integrated boost technique for early-stage breast cancer: A multivariable prediction model. *Radiother. Oncol.*, 122:45–49 (pp. 59, 70, 71).

Harlan, L. C., Clegg, L. X., Abrams, J., Stevens, J. L., & Ballard-Barbash, R. 2006. Community-based use of chemotherapy and hormonal therapy for early-stage breast cancer: 1987–2000. *J. Clin. Oncol.*, 24:872–877 (p. 79).

Harrell, F. E. 2001, Regression modeling strategies, with applications to linear models, survival analysis and logistic regression (New York: Springer) (pp. 9, 17).

Harrell, F. E., Lee, K. L., & Mark, D. B. 1996. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.*, 15:361–387 (pp. 9, 12).

Harris, J. R., Levene, M. B., Svensson, G., & Hellman, S. 1979. Analysis of cosmetic results following primary radiation therapy for stages I and II carcinoma of the breast. *Int. J. Radiat. Oncol. Biol. Phys.*, 5:257–261 (pp. 44, 61).

Hau, E., Browne, L. H., Khanna, S., et al. 2012. Radiotherapy breast boost with reduced whole-breast dose is associated with improved cosmesis: The results of a comprehensive assessment from the St. George and Wollongong randomized breast boost trial. *Int. J. Radiat. Oncol. Biol. Phys.*, 82:682–689 (p. 52).

Haviland, J. S., Hopwood, P., Mills, J., et al. 2016. Do patient-reported outcome measures agree with clinical and photographic assessments of normal tissue effects after breast radiotherapy? The experience of the standardisation of breast radiotherapy (START) trials in early breast cancer. *Clin. Oncol.*, 28:345–353 (pp. 43, 49, 52).

Haviland, J. S., Sydenham, M., Mills, J., et al. 2012. Can patient reported outcome measures replace clinical assessments in breast radiotherapy trials? *Radiother. Oncol.*, 103:S53 (p. 52).

Haviland, J. S., Owen, J. R., Dewar, J. A., et al. 2013. The UK standardisation of breast radiotherapy (START) trials of radiotherapy hypofractionation for treatment of early breast cancer: 10-year follow-up results of two randomised controlled trials. *Lancet Oncol.*, 14:1086–1094 (pp. 70, 72).

Van der Heiden-van der Loo, M., de Munck, L., Visser, O., et al. 2012. Variation between hospitals in surgical margins after first breast-conserving surgery in the Netherlands. *Breast Cancer Res. Treat.*, 131:691–698 (p. 91).

Van der Heiden-van der Loo, M., Siesling, S., Wouters, M. W. J. M., et al. 2015. The value of ipsilateral breast tumor recurrence as a quality indicator: Hospital variation in the Netherlands. *Annals Surg. Oncol.*, 22 Suppl 3:S522–S528 (pp. 73, 91).

Heil, J., Dahlkamp, J., Golatta, M., et al. 2011. Aesthetics in breast conserving therapy: Do objectively measured results match patients' evaluations? *Annals Surg. Oncol.*, 18:134–138 (p. 51).

Hijal, T., Fournier-Bidoz, N., Castro-Pena, P., et al. 2010. Simultaneous integrated boost in breast conserving treatment of breast cancer: A dosimetric comparison of helical tomotherapy and three-dimensional conformal radiotherapy. *Radiother. Oncol.*, 94:300–306 (p. 18).

Holland, R., Peterse, J. L., Millis, R. R., et al. 1994. Ductal carcinoma in situ: A proposal for a new classification. *Semin. Diagn. Pathol.*, 11:167–180 (p. 25).

Houssami, N., & Morrow, M. 2014. Margins in breast conservation: A clinician's perspective and what the literature tells us. *J. Surg. Oncol.*, 110:2–7 (p. 91).

Hughes, K. S., Schnaper, L. A., Berry, D., et al. 2004. Lumpectomy plus tamoxifen with or without irradiation in women 70 years of age or older with early breast cancer. *New Engl. J. Med.*, 351:971–977 (p. 32).

Hughes, K. S., Schnaper, L. A., Bellon, J. R., et al. 2013. Lumpectomy plus tamoxifen with or without irradiation in women age 70 years or older with early breast cancer: Long-term follow-up of CALGB 9343. *J. Clin. Oncol.*, 31:2382–2387 (pp. 32, 92).

Iasonos, A., Schrag, D., Raj, G. V., & Panageas, K. S. 2008. How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.*, 26:1364–1370 (p. 15).

Immink, J. M., Putter, H., Bartelink, H., et al. 2012. Long-term cosmetic changes after breast-conserving treatment of patients with stage I–II breast cancer and included in the EORTC 'boost versus no boost' trial. *Annals Oncol.*, 23:2591–2598 (pp. 59, 68, 69, 70, 71).

Jones, H. A., Antonini, N., Hart, A. A. M., et al. 2009. Impact of pathological characteristics on local relapse after breast-conserving therapy: A subgroup analysis of the EORTC boost versus no boost trial. *J. Clin. Oncol.*, 27:4939–4947 (pp. 7, 14, 24, 27, 32, 91).

Kattan, M. W. 2011. Doc, What are my chances? A conversation about prognostic uncertainty. *Eur. Urol.*, 59:224 (p. 137).

Kehoe, S., Hook, J., Nankivell, M., et al. 2015. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): An open-label, randomised, controlled, non-inferiority trial. *Lancet*, 386:249–257 (p. 117).

Kelemen, G., Varga, Z., Lázár, G., Thurzó, L., & Kahán, Z. 2012. Cosmetic outcome 1–5 years after breast conservative surgery, irradiation and systemic therapy. *Pathol. & Oncol. Res.*, 18:421–427 (p. 52).

Keller, L. M. M., Sopka, D. M., Li, T., et al. 2012. Five-year results of whole breast intensity modulated radiation therapy for the treatment of early stage breast cancer: The Fox Chase Cancer Center experience. *Int. J. Radiat. Oncol. Biol. Phys.*, 84:881–887 (p. 71).

Kirova, Y. M., Castro Pena, P., Hijal, T., et al. 2010. Improving the definition of tumor bed boost with the use of surgical clips and image registration in breast cancer patients. *Int. J. Radiat. Oncol. Biol. Phys.*, 78:1352–1355 (p. 73).

Kreike, B., Halfwerk, H., Armstrong, N., et al. 2009. Local recurrence after breast-conserving therapy in relation to gene expression patterns in a large series of patients. *Clin. Cancer Res.*, 15:4181–4190 (p. 7).

Kunkler, I. H., Williams, L. J., Jack, W. J. L., et al. 2015. Breast-conserving surgery with or without irradiation in women aged 65 years or older with early breast cancer (PRIME II): A randomised controlled trial. *Lancet Oncol.*, 16:266–273 (p. 32).

Van der Laan, H. P., Hurkmans, C. W., Kuten, A., Westenberg, H. A., & on behalf of the EORTC-ROG Breast Working Party. 2010. Current technological clinical practice in breast radiotherapy; results of a survey in EORTC-Radiation Oncology Group affiliated institutions. *Radiother. Oncol.*, 94:280–285 (p. 18).

Van Laar, C., van der Sangen, M. J. C., Poortmans, P. M. P., et al. 2013. Local recurrence following breast-conserving treatment in women aged 40 years or younger: Trends in risk and the impact on prognosis in a population-based cohort of 1143 patients. *Eur. J. Cancer*, 49:3093–3101 (pp. 90, 92).

Lansu, J. T. P., Essers, M., Voogd, A. C., et al. 2015. The influence of simultaneous integrated boost, hypofractionation and oncoplastic surgery on cosmetic outcome and PROMs after breast conserving therapy. *Eur. J. Surg. Oncol.*, 41:1411–1416 (pp. 71, 73).

Lasch, F., Weber, K., Chao, M. M., & Koch, A. 2017. A plea to provide best evidence in trials under sample-size restrictions: The example of pioglitazone to resolve leukoplakia and erythroplakia in Fanconi anemia patients. *Orphanet J. Rare Dis.*, 12 (p. 2).

Lazzari, G., Terlizzi, A., Della Vittoria Scarpati, G., et al. 2017. Predictive parameters in hypofractionated whole-breast 3D conformal radiotherapy according to the Ontario Canadian trial. *Onco Targets Ther.*, 10:1835–1842 (p. 70).

Van der Leest, M., Evers, L., van der Sangen, M. J. C., et al. 2007. The safety of breast-conserving therapy in patients with breast cancer aged $\leq$ 40 years. *Cancer*, 109:1957–1964 (pp. 79, 90, 92).

Litière, S., Werutsky, G., Fentiman, I. S., et al. 2012. Breast conserving therapy versus mastectomy for stage I–II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. *Lancet Oncol.*, 13:412–419 (pp. 24, 79, 81).

Liu, F.-F., Shi, W., Done, S. J., et al. 2015. Identification of a low-risk luminal a breast cancer cohort that may not benefit from breast radiotherapy. *J. Clin. Oncol.*, 33:2035–2040 (pp. 24, 32).

Lopez-Girona, A., Heintel, D., Zhang, L.-H., et al. 2011. Lenalidomide downregulates the cell survival factor, interferon regulatory factor-4, providing a potential mechanistic link for predicting response. *Br. J. Hematol.*, 154:325–336 (p. 95).

Louwman, W. J., Voogd, A. C., van Dijck, J. A. A. M., et al. 2008. On the rising trends of incidence and prognosis for breast cancer patients diagnosed 1975-2004: A long-term population-based study in southeastern Netherlands. *Cancer Causes & Control.*, 19:97–106 (p. 79).

Lowery, A. J., Kell, M. R., Glynn, R. W., Kerin, M. J., & Sweeney, K. J. 2012. Locoregional recurrence after breast cancer surgery: A systematic review by receptor phenotype. *Breast Cancer Res. Treat.*, 133:831–841 (pp. 24, 32).

Makar, A. P., Tropé, C. G., Tummers, P., Denys, H., & Vandecasteele, K. 2016. Advanced ovarian cancer: Primary or interval debulking? Five categories of patients in view of the results of randomized trials and tumor biology: primary debulking surgery and interval debulking surgery for advanced ovarian cancer. *Oncol.*, 21:745–754 (p. 117).

Mannino, M., & Yarnold, J. R. 2009. Local relapse rates are falling after breast conserving surgery and systemic therapy for early breast cancer: Can radiotherapy ever be safely withheld? *Radiother. Oncol.*, 90:14–22 (p. 18).

McCahill, L. E., Single, R. M., Aiello Bowles, E. J., et al. 2012. Variability in reexcision following breast conservation surgery. *JAMA*, 307:467–475 (p. 91).

Van Meurs, H. S., Tajik, P., Hof, M. H. P., et al. 2013. Which patients benefit most from primary surgery or neoadjuvant chemotherapy in stage IIIC or IV ovarian cancer? An exploratory analysis of the European Organisation for Research and Treatment of Cancer 55971 randomised trial. *Eur. J. Cancer*, 49:3191–3201 (pp. 117, 118, 119, 124, 128).

Meyer, L. A., Cronin, A. M., Sun, C. C., et al. 2016. Use and Effectiveness of Neoadjuvant Chemotherapy for Treatment of Ovarian Cancer. *J. Clin. Oncol.*, 34:3854–3863 (p. 125).

Mo, Q., Gonen, M., & Heller, G. 2010, CPE: Concordance Probability Estimates in survival analysis, R package version 1.4.1 (p. 9).

Mook, S., Van 't Veer, L. J., Rutgers, E. J., et al. 2011. Independent prognostic value of screen detection in invasive breast cancer. *J. Natl. Cancer Inst.*, 103:585–597 (p. 91).

Moran, M. S., Schnitt, S. J., Giuliano, A. E., et al. 2014. Society of Surgical Oncology–American Society for Radiation Oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages I and II invasive breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, 88:553–564 (pp. 24, 32, 91).

Morrison, J., Swanton, A., Collins, S., & Kehoe, S. 2007. Chemotherapy versus surgery for initial treatment in advanced ovarian epithelial cancer. *Cochrane Database Syst. Rev.*:CD005343 (p. 117).

Morrow, M. 2008. Margins in breast-conserving therapy: Have we lost sight of the big picture? *Expert. Rev. Anticancer. Ther.*, 8:1193–1196 (pp. 7, 17).

Morrow, M., Jagsi, R., Alderman, A. K., et al. 2009. Surgeon recommendations and receipt of mastectomy for treatment of breast cancer. *JAMA*, 302:1551–1556 (p. 91).

Morschhauser, F., Feugier, P., Flinn, I. W., et al. 2021. A phase 2 study of venetoclax plus R-CHOP as first-line treatment for patients with diffuse large B-cell lymphoma. *Blood*, 137:600–609 (p. 95).

Mukesh, M. B., Harris, E., Collette, S., et al. 2013a. Normal tissue complication probability (NTCP) parameters for breast fibrosis: Pooled results from two randomised trials. *Radiother. Oncol.*, 108:293–298 (pp. 59, 70).

Mukesh, M. B., Barnett, G. C., Wilkinson, J. S., et al. 2013b. Randomized controlled trial of intensity-modulated radiotherapy for early breast cancer: 5-year results confirm superior overall cosmesis. *J. Clin. Oncol.*, 31:4488–4495 (pp. 43, 70).

Mukesh, M. B., Qian, W., Wilkinson, J. S., et al. 2014. Patient reported outcome measures (PROMs) following forward planned field-in field IMRT: Results from the Cambridge Breast IMRT trial. *Radiother. Oncol.*, 111:270–275 (pp. 43, 50, 52).

NABON. 2012, Richtlijn 'Behandeling van het Mammacarcinoom', Tech. rep., Kwaliteitsinstituut voor de Gezondheidszorg CBO/Vereniging van de Integrale Kankercentra en Nationaal Borstkanker Overleg Nederland (p. 80).

Offersen, B. V., Overgaard, M., Kroman, N., & Overgaard, J. 2009. Accelerated partial breast irradiation as part of breast conserving therapy of early breast carcinoma: A systematic review. *Radiother. Oncol.*, 90:1–13 (p. 18).

Oki, Y., Noorani, M., Lin, P., et al. 2014. Double hit lymphoma: The MD Anderson Cancer Center clinical experience. *Br. J. Haematol.*, 166:891–901 (p. 95).

Ozenne, B. M. H., Scheike, T. H., Staerk, L., & Gerds, T. A. 2020. On the estimation of average treatment effects with right-censored time to event outcome and competing risks. *Biom. J.*, 62:751–763 (p. 98).

Perez, E. A., Romond, E. H., Suman, V. J., et al. 2014. Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: Planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. *J. Clin. Oncol.*, 32:3744–3752 (p. 79).

Peterson, D., Truong, P. T., Parpia, S., et al. 2015. Predictors of adverse cosmetic outcome in the RAPID trial: An exploratory analysis. *Int. J. Radiat. Oncol. Biol. Phys.*, 91:968–976 (pp. 59, 70, 71).

Peterson, M. E., Schultz, D. J., Reynolds, C., & Solin, L. J. 1999. Outcomes in breast cancer patients relative to margin status after treatment with breast-conserving surgery and radiation therapy: The University of Pennsylvania experience. *Int. J. Radiat. Oncol. Biol. Phys.*, 43:1029–1035 (p. 17).

Piantadosi, S. 2017, Clinical Trials: A Methodologic Perspective (John Wiley & Sons) (pp. 2, 3).

Polgár, C., Van Limbergen, E., Pötter, R., et al. 2010. Patient selection for accelerated partial-breast irradiation (APBI) after breast-conserving surgery: Recommendations of the Groupe Européen de Curiethérapie-European Society for Therapeutic Radiology and Oncology (GEC-ESTRO) breast cancer working group based on clinical evidence (2009). *Radiother. Oncol.*, 94:264–273 (p. 18).

Poortmans, P., Aznar, M., & Bartelink, H. 2012. Quality indicators for breast cancer: Revisiting historical evidence in the context of technology changes. *Semin. Radiat. Oncol.*, 22:29–39 (pp. 32, 79).

Poortmans, P. M., Collette, L., Bartelink, H., et al. 2008. The addition of a boost dose on the primary tumour bed after lumpectomy in breast conserving treatment for breast cancer. A summary of the results of EORTC 22881-10882 'boost versus no boost' trial. *Cancer/Radiotherapie*, 12:565–570 (pp. 59, 81).

Poortmans, P. M., Collette, L., Horiot, J.-C., et al. 2009. Impact of the boost dose of 10 Gy versus 26 Gy in patients with early stage breast cancer after a microscopically incomplete lumpectomy: 10-year results of the randomised EORTC boost trial. *Radiother. Oncol.*, 90:80–85 (p. 7).

Poortmans, P. M. P., Arenas, M., & Livi, L. 2017. Over-irradiation. *Breast*, 31:295–302 (p. 73).

Pritchard, K. I., & Sousa, B. 2011. Long-term follow-up of women in trials of adjuvant therapy for breast cancer: Is it still important? *J. Clin. Oncol.*, 29:1651–1652 (p. 32).

R Core Team. 2010, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. (p. 9).

Ribelles, N., Perez-Villa, L., Jerez, J. M., et al. 2013. Pattern of recurrence of early breast cancer is different according to intrinsic subtype and proliferation index. *Breast Cancer Res.*, 15:R98 (p. 32).

Rosenwald, A., Bens, S., Advani, R., et al. 2019. Prognostic significance of *MYC* rearrangement and translocation partner in diffuse large B-cell lymphoma: A study by the lunenburg lymphoma biomarker consortium. *J. Clin. Oncol.*, 37:3359–3368 (p. 95).

Rudloff, U., Jacks, L. M., Goldberg, J. I., et al. 2010. Nomogram for predicting the risk of local recurrence after breast-conserving surgery for ductal carcinoma in situ. *J. Clin. Oncol.*, 28:3762–3769 (p. 16).

Rutten, M. J., Leeflang, M. M. G., Kenter, G. G., Mol, B. W. J., & Buist, M. 2014. Laparoscopy for diagnosing resectability of disease in patients with advanced ovarian cancer. *Cochrane Database Syst. Rev.*:CD009786 (p. 117).

Rutten, M. J., van de Vrie, R., Bruining, A., et al. 2015. Predicting surgical outcome in patients with International Federation of Gynecology and Obstetrics stage III or IV ovarian cancer using computed tomography: A systematic review of prediction models. *Int. J. Gynecol. Cancer*, 25:407–415 (pp. 117, 127).

Rutten, M. J., van Meurs, H. S., van de Vrie, R., et al. 2017. Laparoscopy to predict the result of primary cytoreductive surgery in patients with advanced ovarian cancer: A randomized controlled trial. *J. Clin. Oncol.*, 35:613–621 (pp. 117, 127).

Sanders, M. E., Scroggins, T., Ampil, F. L., & Li, B. D. 2007. Accelerated partial breast irradiation in early-stage breast cancer. *J. Clin. Oncol.*, 25:996–1002 (p. 92).

Sanghani, M., Balk, E., Cady, B., & Wazer, D. 2007. Predicting the risk of local recurrence in patients with breast cancer: An approach to a new computer-based predictive tool. *Am. J. Clin. Oncol.*, 30:473–480 (pp. 7, 16).

Sanghani, M., Truong, P. T., Raad, R. A., et al. 2010. Validation of a web-based predictive nomogram for ipsilateral breast tumor recurrence after breast conserving therapy. *J. Clin. Oncol.*, 28:718–722 (p. 7).

Savage, K. J., Johnson, N. A., Ben-Neriah, S., et al. 2009. *MYC* gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood*, 114:3533–3537 (p. 95).

Schemper, M. 1992. Cox analysis of survival data with non-proportional hazard functions. *J. Royal Stat. Soc. Ser. D*, 41:455–465 (p. 97).

Scott, N. W., McPherson, G. C., Ramsay, C. R., & Campbell, M. K. 2002. The method of minimization for allocation to clinical trials. A review. *Control. Clin. Trials*, 23:662–674 (p. 60).

Sehn, L. H., Berry, B., Chhanabhai, M., et al. 2007. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood*, 109:1857–1861 (p. 95).

Siegel, R., Naishadham, D., & Jemal, A. 2012. Cancer statistics, 2012. *CA: Cancer J. for Clin.*, 62:10–29 (p. 117).

Sneeuw, K. C., Aaronson, N. K., Yarnold, J. R., et al. 1992. Cosmetic and functional outcomes of breast conserving treatment for early stage breast cancer. 1. Comparison of patients' ratings, observers' ratings, and objective assessments. *Radiother. Oncol.*, 25:153–159 (pp. 45, 48, 51, 61).

Sobin, L. H., , Wittekind, C., & International Union against Cancer. 2002, TNM Classification of Malignant Tumors, 6th edn. (New York: Wiley-Liss) (pp. 43, 60).

Steyerberg, E. W. 2009, Clinical prediction models: A practical approach to development, validation, and updating (New York: Springer) (p. 16).

Sukel, M. P. P., van de Poll-Franse, L. V., Nieuwenhuijzen, G. A. P., et al. 2008. Substantial increase in the use of adjuvant systemic treatment for early stage breast cancer reflects changes in guidelines in the period 1990–2006 in the southeastern Netherlands. *Eur. J. Cancer*, 44:1846–1854 (p. 79).

Swerdlow, S. H., Campo, E., Pileri, S. A., et al. 2016. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood*, 127:2375–2390 (p. 95).

The Catalogue of Bias Collaboration. 2022, Catalogue of Bias, `https://catalogofbias.org/`, accessed: 14 July 2022 (p. 1).

The International Non-Hodgkin's Lymphoma Prognostic Factors Project. 1993. A predictive model for aggressive non-Hodgkin's lymphoma. *New Engl. J. Med.*, 329:987–994 (p. 95).

The START Trialists' Group. 2008. The UK Standardisation of breast radiotherapy (START) Trial A of radiotherapy hypofractionation for treatment of early breast cancer: A randomised trial. *Lancet Oncol.*, 9:331–341 (p. 18).

Therneau, T. M., & Grambsch, P. M. 2000, Modeling survival data: Extending the Cox model (New York: Springer) (p. 26).

Tramm, T., Mohammed, H., Myhre, S., et al. 2014. Development and validation of a gene profile predicting benefit of postmastectomy radiotherapy in patients with high-risk breast cancer: A study of gene expression in the DBCG82bc cohort. *Clin. Cancer Res.*, 20:5272–5280 (p. 33).

U.S. Department of Health and Human Services Food and Drug Administration. 2018, Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics Guidance for Industry (p. 2).

Veldeman, L., Schiettecatte, K., De Sutter, C., et al. 2016. The 2-year cosmetic outcome of a randomized trial comparing prone and supine whole-breast irradiation in large-breasted women. *Int. J. Radiat. Oncol. Biol. Phys.*, 95:1210–1217 (p. 70).

Vergote, I., De Wever, I., Tjalma, W., et al. 1998. Neoadjuvant chemotherapy or primary debulking surgery in advanced ovarian carcinoma: A retrospective analysis of 285 patients. *Gynecol. Oncol.*, 71:431–436 (p. 117).

Vergote, I., Tropé, C. G., Amant, F., et al. 2011. Neoadjuvant chemotherapy is the better treatment option in some patients with stage IIIC to IV ovarian cancer. *J. Clin. Oncol.*, 29:4076–4078 (p. 117).

—. 2010. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. *New Engl. J. Med.*, 363:943–953 (pp. 117, 119, 127).

Vergote, I. B., Van Nieuwenhuysen, E., & Vanderstichele, A. 2016. How to select neoadjuvant chemotherapy or primary debulking surgery in patients with stage IIIc or IV ovarian carcinoma. *J. Clin. Oncol.*, 34:3827–3828 (p. 117).

Veronesi, U., Cascinelli, N., Mariani, L., et al. 2002. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *New Engl. J. Med.*, 347:1227–1232 (p. 79).

Voduc, K. D., Cheang, M. C. U., Tyldesley, S., et al. 2010. Breast cancer subtypes and the risk of local and regional relapse. *J. Clin. Oncol.*, 28:1684–1691 (p. 32).

Voogd, A. C., Nielsen, M., Peterse, J. L., et al. 2001. Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II breast cancer: Pooled results of two large European randomized trials. *J. Clin. Oncol.*, 19:1688–1697 (p. 24).

Vrieling, C., Collette, L., Fourquet, A., et al. 1999. The influence of the boost in breast-conserving therapy on cosmetic outcome in the EORTC "boost versus no boost" trial. *Int. J. Radiat. Oncol. Biol. Phys.*, 45:677–685 (pp. 43, 52, 59, 68, 70, 71).

—. 2003. Can patient-, treatment- and pathology-related characteristics explain the high local recurrence rate following breast-conserving therapy in young patients? *Eur. J. Cancer*, 39:932–944 (pp. 24, 90).

Wapnir, I. L., Anderson, S. J., Mamounas, E. P., et al. 2006. Prognosis after ipsilateral breast tumor recurrence and locoregional recurrences in five national surgical adjuvant breast and bowel project node-positive adjuvant breast cancer trials. *J. Clin. Oncol.*, 24:2028–2037 (p. 91).

Whelan, T. J., Pignol, J.-P., Levine, M. N., et al. 2010. Long-term results of hypofractionated radiation therapy for breast cancer. *New Engl. J. Med.*, 362:513–520 (pp. 18, 70).

Williamson, D., Dinniwell, R., Fung, S., et al. 2010. Local control with conventional and hypofractionated adjuvant radiotherapy after breast-conserving surgery for ductal carcinoma in-situ. *Radiother. Oncol.*, 95:317–320 (p. 18).

Wright, A. A., Bohlke, K., Armstrong, D. K., et al. 2016. Neoadjuvant chemotherapy for newly diagnosed, advanced ovarian cancer: Society of Gynecologic Oncology and American Society of Clinical Oncology clinical practice guideline. *Gynecol. Oncol.*, 143:3–15 (p. 117).

Yıldırım, M., Kaya, V., Özlem Demirpençe, & Paydaş, S. 2015. The role of gender in patients with diffuse large B cell lymphoma treated with rituximab-containing regimens: A meta-analysis. *Arch. Med. Sci.*, 11:708–714 (p. 95).

Young, J., Badgery-Parker, T., Dobbins, T., et al. 2015. Comparison of ECOG/WHO performance status and ASA score as a measure of functional status. *J. Pain Symptom Manag.*, 49:258–264 (p. 43).

Yu, T., Eom, K.-Y., Jang, N. Y., et al. 2016. Objective measurement of cosmetic outcomes of breast conserving therapy using BCCT.core. *Cancer Res. Treat.*, 48:491–498 (p. 51).