Statistical methods to support medical decision-making on patient positioning for breast irradiation

Ph.D. Thesis

Doctoral School of Interdisciplinary Medicine, University of Szeged

Ferenc Rárosi, M.Sc.

Department of Medical Physics and Informatics, University of Szeged

Supervisors:

Dr. Krisztina Boda, Ph.D.

Dr. Tibor Nyári, Ph.D., D.Sc.

Department of Medical Physics and Informatics, University of Szeged

Szeged, 2020

“An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem.”

– John Tukey

“To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.”

– Sir Ronald Fisher

The thesis is based on the following publications by the author

I. **Rárosi, F.**, Boda, K., Kahán, Z., and Varga, Z. (2019). Decision curve analysis apropos of choice of preferable treatment positioning during breast irradiation. BMC Medical Informatics and Decision Making 19(1), 204. **IF: 2.067**

II. Varga, Z., Cserháti, A., **Rárosi, F.**, Boda, K., Gulyás, G., Együd, Z., and Kahán, Z. (2014). Individualized positioning for maximum heart protection during breast irradiation. Acta Oncologica 53, 58–64. **IF: 2.997**

III. Kahán, Z., **Rárosi, F.**, Gaál, S., Cserháti, A., Boda, K., Darázs, B., Kószó, R., Lakosi ,F., Gulybán, Á., and Varga, Z. (2018). A simple clinical method for predicting the benefit of prone vs. supine positioning in reducing heart exposure during left breast radiotherapy. Radiotherapy and Oncology 126(3), 487–492. **IF: 4.942**

Abbreviations

AUC_{ROC}	area under ROC curve
BS	Brier score
CDF	cumulative distribution function
CI	confidence interval (typically 95% confidence level)
CT	computed tomography
ΔMD_{LAD}	LAD mean dose difference (primary dependent variable in this thesis)
DCA	decision curve analysis
DIBH	deep inspiration breath hold
DVH	dose volume histogram
EDF	empirical cumulative distribution function
FN	number of false negative cases
FP	number of false positive cases
Gy	Gray (absorbed dose measure 1 Gy=1 J/kg)
IMRT	intensity-modulated radiotherapy
LAD	left anterior descendent (coronary artery)
LR	likelihood ratio
NB	net benefit
OAR	organ at risk
OLS	ordinary least squares (estimation method)
P_t	threshold probability
PTV	planning target volume
R	Pearson’s product moment correlation coefficient

R^2	coefficient of determination (square of the Pearson correlation)
R^2_{adj}	squared Pearson correlation coefficient adjusted for degrees of freedom
ROC	receiver operating characteristic
TN	number of true negative cases
TP	number of true positive cases
V_{25Gy}	percentage volume of the heart receiving more than 25 Gy

1. Introduction

Mathematical statistics is very important in all fields of empirical science. Biostatistics is the application of mathematical statistical methods in biomedical research.

Prediction models are widely used in biomedical research and other interdisciplinary fields of research, such as economics and engineering. If a dependent variable is continuous, then a multiple regression model can be used, while logistic regression is employed for categorical dependent variables. The result of a regression model is an expected value of the dependent variable, and the result of a (binary) logistic regression is an expected probability [1]. When the purpose is to make a decision and make predictions concerning the existence of a phenomenon, such as an illness or the necessity of an operation, the decision is based on a carefully chosen cut-point. The performance of a prediction model can be evaluated by comparing the decision to the gold standard method. In most cases, we simply do not know the 'truth'. The nearest to the truth we have is the gold standard, so we have to regard that as the 'truth'.

There are several measures to describe the performance of the prediction model based on the numbers in true positive (TP), false positive (FP), true negative (TN) and false negative (FN) cases. The well-known measures are sensitivity, specificity, positive predictive value, negative predictive value, accuracy (proportion of all correct diagnoses) and the Youden index (i.e. sensitivity+specificity-1). The method, which involves ROC curves, is based on the measures noted above: we plot the sensitivity in function of 1-specificity at various threshold levels.

ROC analysis is the most commonly used method for measuring the diagnostic ability of a prediction method or diagnostic test. The ROC method is very effective, as the performance of the model can be measured with one number (i.e. the area under the ROC curve). However, it has the disadvantage of equally weighing false positive and false negative decisions, whereas the weighting of the different types of misclassifications may be important.

Andrew J. Vickers and Elena B. Elkin published a new method which not only eliminates this weakness in the ROC method, but also introduces a completely new approach to derive a measure called 'net benefit' to evaluate the clinical utility of the method [2]. The value of the net benefit depends on the cut-point chosen. We can obtain the decision curve if we plot the net benefit in function of the cut-point (threshold probability). Decision curves can be effectively used to compare different prediction methods and to determine the range of the possible cut-point, where the use of the prediction model is beneficial.

In this study, we apply the decision curve method to evaluate and compare our models for the breast irradiation positioning problem.

1.1. The breast irradiation positioning problem

Radiotherapy is an effective treatment for breast cancer, but it can lead to significant late morbidity, particularly that related to heart damage caused by radiogenic heart diseases. The goal of radiotherapy is to irradiate the volume specified as the target volume (volume to be irradiated) with the prescribed (planned) dose. Also, normal tissues and especially organs at risk (the organs and tissues to be spared) should be exposed with radiation doses that are as low as possible. Our goal was to minimize the irradiation dose to the left anterior descendent (LAD) coronary artery, the LAD being an organ at risk. Later, we will simply call this dose value the LAD dose. The radiotherapy can be performed in a supine or prone position. We consider the prone position preferable if the LAD dose is smaller in that position than in the supine position, and we call the supine position preferable if the opposite is true. The preferable position varies from patient to patient.

A series of CT scans and therapy plans are needed in both positions (supine and prone) for the gold standard decision on the preferable treatment position. This method is expensive (with respect to the technology and work of physicians) and involves an extra dose of radiation for patients. That was the motivation for creating a model to predict the preferable position and anticipate the dose difference between the two positions using patients' characteristics.

1.2. Aims

The aims of our interdisciplinary research project were the following:

(1) To analyse the relationship between the dependent variable (LAD mean dose difference) and possible predictor variables and to select possible predictor variables for model building.

(2) To develop a model-based classifier method to predict the preferable treatment position.

(3) To evaluate the performance of the prediction models, to apply a novel method involving decision curves to compare different prediction models (model selection), and to carry out simulations to confirm evaluation of our models with decision curves.

(4) To validate the final model with random cross-validation and further test the model on an independent dataset of patients.

(5) To investigate the possible effects of predictor measurement errors (plane miss).

(6) To apply the Kolmogorov–Smirnov method to construct a 95% confidence band for LAD dose distribution and to determine dose constraints based on LAD dose distribution data.

In addition, this study contains what are hoped to be interesting methods to predict the preferable treatment position and offers ideas for further research.

Finally, the application of mathematical statistics might be a useful tool in radiation treatment planning for left-side breast cancer patients for physicians and physicists. It is anticipated that this thesis will aid in individual treatment positioning to minimize the heart dose of left-side breast cancer radiotherapy.

2. Methods

This section deals with the commonly used prediction methods employed in this study. A short description of the well-known classical methods is provided to facilitate comprehension among readers less versed in statistics. Since the methods used in the thesis are known, it is clear that a detailed methodological explanation of well-established methods should not be a goal of this thesis. However, a few methods to evaluate the performance of a prediction model will be described, and the relatively newer method of decision curves will be explained in greater detail, especially decision curves for 'non-probability outcomes'. The structure of the dataset will also be covered after the methodological introduction.

2.1. Commonly used simple prediction models

There are many well-known classification methods in mathematical statistics used in this thesis. A detailed description of these methods is available in numerous statistics books. Thus, these methods are just briefly mentioned in the thesis. However, to demonstrate the application of these methods in the field of radiotherapy and in the development of new prediction models might be regarded as a result in itself. It is important to emphasize that these methods are not self-developed, but their careful adaptation in the field of individualized radiotherapy can be considered as an original contribution.

2.1.1. Linear regression

If the dependent (response) variable *y* is continuous, then linear regression can be used. The outcome of the linear regression approximates the dependent variable.

$$\hat{y} = a_1 x_1 + \cdots + a_k x_k + b, \quad (1)$$

where $a_1 \ldots a_k$ are the regression coefficients for predictors $x_1 \ldots x_k$, respectively, and b is a constant.

Assumptions for the linear regression model are

$$E(y|x) = Y(x) = f(x, \alpha_1 \ldots, \alpha_k, \beta), \quad (2)$$

where f is the known type of function and $\alpha_1 \ldots \alpha_k, \beta$ are parameters of the function (and $a_1 \ldots a_k$, b are estimates for the real, unknown parameters $\alpha_1 \ldots \alpha_k, \beta$). In the simplest case (i.e. a linear relationship), we assume that random variable y is the sum of the linear combination of the predictors, a constant and an error term.

$$y = \alpha_1 x_1 + \cdots + \alpha_k x_k + \beta + \epsilon \quad (3)$$

$$Var(y) = Var(y|x_{1\ldots}x_k) = constant, \quad (4)$$

where the conditional variance of *y* is constant (it does not depend on the predictor values) or a known function of $x_1 \ldots x_k$,

the residuals are independent of each other, and

$$\varepsilon \sim N(0, \sigma^2), \quad (5)$$

the residuals follow a normal distribution.

A very detailed description of the linear regression method is available in a range of statistics books [3-6].

2.1.2. Logistic regression

If the dependent (response) variable (Y) is binary (typically diseased and control, coded as Y=1 and Y=0, respectively), then a binary logistic regression can be used. The outcome of the binary logistic regression is an expected probability: $\hat{P} = P(Y = 1|x_1 \ldots x_k)$.

$$logit(\hat{P}) = ln\frac{\hat{P}}{1-\hat{P}} = a_1x_1 + \cdots + a_kx_k + b, \quad (6)$$

since

$$\hat{P} = \frac{e^{a_1x_1+\cdots+a_kx_k+b}}{1+e^{a_1x_1+\cdots+a_kx_k+b}} = \frac{1}{1+e^{-(a_1x_1+\cdots+a_kx_k+b)}}, \quad (7)$$

where $\hat{P}$ is the expected probability of the disease, $a_1 \ldots a_k$ are the estimated coefficients for predictors $x_1 \ldots x_k$, respectively, and *b* is the constant.

Assumptions for the logistic regression may not be regarded as being as strict as those for the linear regression, but still there are plausible requirements [7].

One requirement is the independence of errors (i.e. no duplicate responses).

Another requirement is the linearity of the logit for continuous predictors (or at least a monotonic relationship between predictor values and the logit of the outcome).

A lack of correlation between the predictors is also required, so relatively poorly correlated predictors should be chosen to avoid multicollinearity.

Influential outliers (i.e. when the parameters/predictors suggest a different group than the real group membership) may reduce the predictive performance of the model.

There is a sample size requirement as well. As a rule of a thumb, at least ten observations are needed per predictor variable to avoid overfit and large standard errors of the coefficient estimates [8].

In this thesis, applying logistic regression-based models is extraordinarily important for predictive performance [9]. A detailed description of logistic regression models can be read in [8, 10].

2.2. Measures to evaluate the predictive power of classification methods

2.2.1. Classical measures

The performance of a prediction model can be evaluated by comparing the decision to the gold standard method. There are several measures to describe the performance of the prediction model based on the numbers in TP, FP, TN and FN cases. The well-known measures are sensitivity, specificity, positive predictive value, negative predictive value, accuracy (the proportion of all correct diagnoses) and the Youden index (i.e. sensitivity+specificity-1).

2.2.2. ROC analysis

The method involving ROC curves is based on the classical measures noted above. Let the horizontal axis of the coordinate system be 1-specificity and the vertical axis be sensitivity. This coordinate system is called the ROC space. In the ROC method, we plot the sensitivity in function of 1-specificity at various threshold levels. The ‘ideal point’ of the ROC space is the point (0,1) because this means 100% sensitivity and 100% specificity for the same cut-point. The two extremal ROC curves are shown in Figure 1. The one on the left refers to the best possible classification, i.e. the perfect classification, since the curve crosses the ‘ideal’ point (0,1). The one on the right indicates random guessing. In the right-hand one, sensitivity equals 1-specificity, which means that the sum of sensitivity and specificity is 1 and the Youden index is a constant 0 for all possible cut-points, which means, for example, 50% sensitivity and 50% specificity, or 100% sensitivity with 0% specificity.

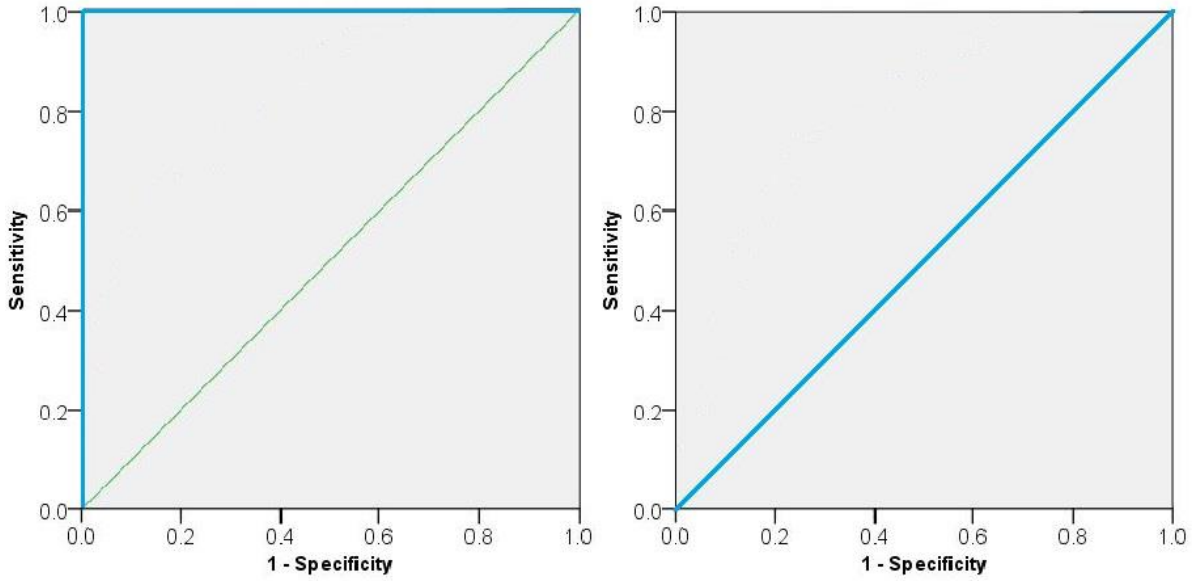


Figure 1. Extremal ROC curves. The left-hand one refers to the perfect classification, and the right-hand one indicates the worst possible classification (i.e. random guessing).

ROC analysis is the most commonly used method for measuring the diagnostic performance of a test. The ROC method can be regarded as an effective method, as the performance of the model can be measured with one number (i.e. the area under the ROC curve). However, it has the disadvantage of equally weighting false positive and false negative decisions, whereas the weighting of the different types of misclassifications may be important.

We can choose the optimal cut-point based on these measures. One possible method for cut-point selection is to find the maximum value of the Youden index, and the value that maximizes the Youden index is the cut-point. There are other methods for choosing a cut-point [10-14].

2.2.3. Brier score

The Brier score, originally introduced by Glenn W. Brier [15], is

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2, \qquad (8)$$

and the weighted Brier score is

$$wBS = \frac{\sum_{i=1}^{N} w_i(p_i - o_i)^2}{\sum_{i=1}^{N} w_i}, \qquad (9)$$

where o_i is the ith outcome, p_i is the ith predicted probability, and w_i is the weight of the ith item.

If the outcome occurs in the ith case, then $o_i=1$; otherwise, $o_i=0$. The Brier score measures an average (or weighted average) of squared distances between the current outcomes and the predicted probabilities, lower values indicating better predictive performance.

2.2.4. Decision curve analysis (DCA)

Decision curve analysis is a novel method to evaluate the performance of diagnostic tests and a prediction model. If the outcome of a prediction model is a predicted probability, decisions are generally based on a carefully chosen cut-point p_t, above which the prediction is positive and below which the prediction is negative. The decision curve is based on the predicted probabilities of statistical models. The decision curve method was introduced by Andrew J. Vickers and Elena B. Elkin. A decision curve is a plot which shows the net benefit calculated at various threshold levels [2].

Decision curves are more frequently used in biomedical studies to compare different prediction models [16-20] or to check model performance when adding a new predictive marker [21, 22].

2.2.4.1. A novel measure: net benefit

Experts simply cannot weight outcomes based on an easily measurable parameter in a large number of decision problems. This problem is especially true in many applications in the field of medicine because the quantitative description of late consequences may be unknown. Still, there is the need to weight the outcomes of decisions. Here we switch to a brief description of the decision theory methods, and we take into account the 'benefit' and 'loss' of the decisions.

The definition of net benefit is based on the 'utility of the prediction of the method', originally defined by Peirce [23] as

$$B = \frac{p \cdot TP - l \cdot FP}{TP+FP+FN+TN} = \frac{p \cdot TP - l \cdot FP}{N}, \qquad (10)$$

where p stands for the 'profit' (average profit) of true positive decisions, l refers to the 'loss' (average loss) of false positive decisions, TP, FP, TN and FN are the number of true positive, false positive, true negative and false negative decisions, respectively, and N is the sample size.

The net benefit is defined as the benefit divided by the profit:

$$NB = \frac{B}{p} = \frac{TP}{N} - \frac{l}{p}\frac{FP}{N} \qquad (11)$$

In other words, the net benefit is the benefit that results from the normalization of the profit. In this aspect, the 'loss-to-profit ratio' is a weighting factor to give weight to the false positive decision compared to one unit of benefit of the true positive decision. It is important to note that

the profit and loss are unknown in most applications and it is impossible to measure them. This is common in the area of medical decisions. One simply cannot measure or numerically anticipate the consequences of the true positive decision (the so-called profit of the operation, for instance) or the consequences of the false positive decision (for example, the loss of an unnecessary operation). That is why Vickers and Elkin suggested calculating this weighting factor (loss-to-profit ratio) as the odds of the threshold probability. Usually, the weights of the four possible outcomes (TP, FP, FN and TN) are not known; still, it is possible to make an acceptable assumption of the 'loss'-to-'profit' ratio.

There is an important assumption by Vickers [2]:

$$\frac{l}{p} := \frac{p_t}{1-p_t}, \tag{12}$$

where p_t is the threshold probability (or cut-point), above which the outcome of a probability prediction model is labelled 'positive' and below which it is labelled 'negative'.

If we accept this assumption, net benefit simplifies to

$$NB = \frac{B}{p} = \frac{TP}{N} - \frac{p_t}{1-p_t}\frac{FP}{N} \tag{13}$$

Decision curves

The decision curve is a curve that illustrates the net benefit in function of the threshold level p_t. Using this method, we can compare the performance of different predictive models to show which model is more beneficial in function of the threshold probability. This method also shows the range of threshold levels where the decision is beneficial.

An example of decision curves

An example of the decision curve can be seen in Figure 2. There are two extremal decisions. One is to 'treat all' (regardless of the cut-off probability), which means we classify everybody as positive and apply the prevention (or the treatment) to everybody. Surprisingly, this decision can lead to net benefit, depending on the incidence of the positive status (see formula 13). Another extremal decision is to 'treat none', which means that we classify everybody as negative (regardless of the cut-off probability). This leads to a constant zero net benefit (see formula 13), and there are no positive classifications; thus, formula 13 amounts to zero. If a prediction method cannot provide better values of net benefit than those two extremities, then the prediction model is useless with respect to clinical utility. We are looking for models which can provide high values of net benefit in a large range of threshold probabilities.

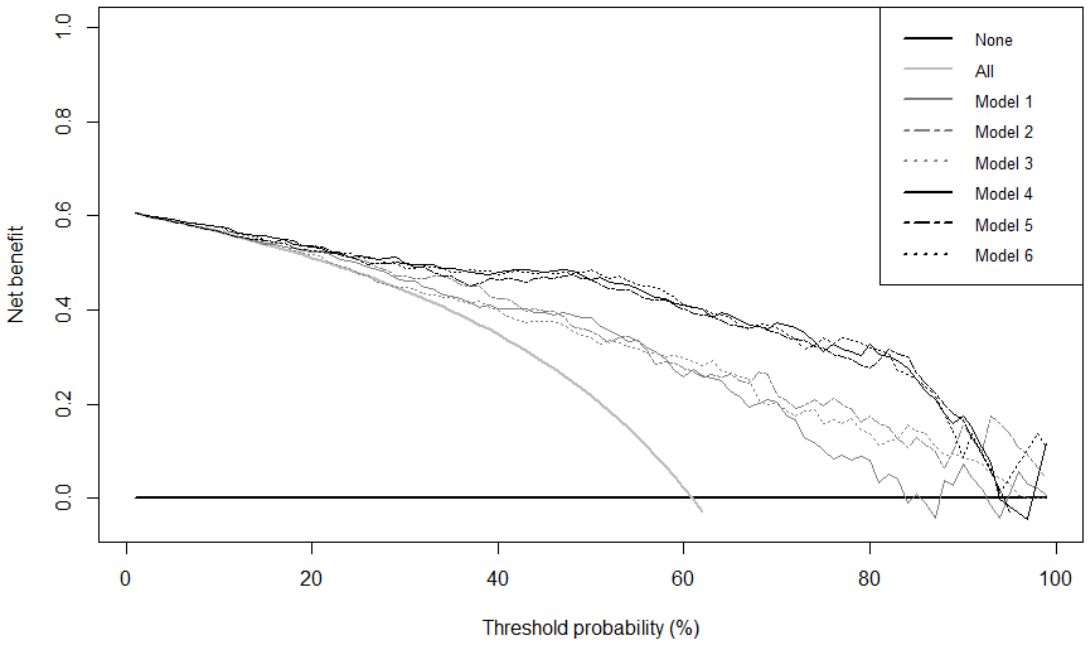


Figure 2. Example of a decision curve. The grey line refers to 'treat all', while the horizontal black line indicates 'treat none'. Models 1 to 3 can offer more net benefit than extremal decisions, but models 4–6 can provide better values of net benefit. Models 4 to 6 are thus preferred.

This example is just meant to provide a graphic understanding. These curves will be presented in the results section, and the models will be explained.

2.2.4.2. Decision curves and 'non-probability outcomes'

In many applications, the result of a prediction model is not a probability estimation, but a score or an expected value (a continuous variable in general), which we call a 'non-probability outcome' in this thesis. The decision curve method is based on predicted probabilities. In order to use DCA for a non-probability outcome, a transformation is needed. We call this transformation a 'link function' in this thesis.

The questionable point of this transformation is that the shape of the decision curve might depend on the transformation function chosen.

In order to examine how the shape of the decision curve depends on the link function chosen, random predictors were simulated from different distributions. In the first approach, two symmetrical and two skewed distributions were generated, and each of them was examined with each of the four transformation methods. In the second approach, two shifted normal distributions were generated with several shift parameters. Again, the effect of the transformations was examined on the shape of the decision curve.

The simulations were carried out in IBM SPSS version 24 and R version 3.3.1, and the detailed results of these investigations can be read in section (3.4).

2.3. On breast radiation treatment positioning (application)

2.3.1. Dataset and candidate predictors

The study was approved by the institutional review board of the University of Szeged (registration number 185/2012), and all the enrolled patients gave their written informed consent to participate. Consecutive patients with left-side breast cancer requiring radiotherapy of the operated breast were included throughout the study.

This interdisciplinary investigation took several years, during which time the sample size accumulated naturally. Our dataset in the first stage consisted of 83 patients with left-side breast cancer at the University of Szeged Department of Oncotherapy. Later, we obtained data for a further 55 patients. I call this the second stage. Most of the calculations were carried out on the first 83 or the first and second stage dataset of 83+55=138 patients [24, 25]. Predictor errors were tested in an additional 100 patients' data. I call this stage 3 [26]. All the data from stages 1 to 3 was derived from the University of Szeged Department of Oncotherapy.

External testing was carried out on data from 28 patients from Liège, which I call stage 4 [26].

The patients' characteristics were age, body weight, body height, BMI (body mass index), 'breast volume' planning target volume (PTV i.e. the operated breast with safety margin), ventral circumference, waist circumference, hip circumference, volume of the lung and volume of the heart. The newer characteristics were: the surface area of the heart at the middle of the LAD, labelled 'area' or 'A_{heart}', and the median distance between the LAD (left anterior descendent coronary artery) and the chest wall, denoted as 'median distance' or 'D_{med}'.

2.3.2. Outcome (dependent variable) LAD mean dose difference

The dependent variable was characteristic of the irradiation dose based on the 'gold standard method'. The 'gold standard method' means a series of CT scans and radiation treatment plans in both positions (the supine and prone positions). We regard this method as the most precise method of decision-making. The primary dependent variable was the difference of the expected dose to the LAD, supine minus prone position, denoted as 'LAD mean dose difference' or 'ΔMD_{LAD}'. If the LAD mean dose difference is positive, then the prone position is regarded as favourable; similarly, negative difference means that the supine position is favourable. The secondary dependent variable

was based on the volume of the heart receiving more than 25 Gy, denoted as V_{25Gy}. The secondary dependent variable was the difference of V_{25Gy} supine minus prone position, denoted as ΔV_{25Gy}. Although quite a few investigations have been conducted involving the secondary outcome variable ΔV_{25Gy}, this thesis focuses on the primary outcome variable and contains the detailed results for the primary outcome variable.

2.3.3. Random cross-validation

The internal validation was done using SPSS macros. In the first step, 1000 Bernoully-distributed variables were generated for the random divisions of the dataset. Regression was run for each training set, and the predictions on each appropriate test set were noted. Then the predicted values were recoded as 0 for correct predictions, 1 for false positives (i.e. the prediction is that supine is better and the gold standard prone is preferable) and -1 for false negatives (i.e. the prediction is that prone is better and the gold standard supine is preferable).

2.3.4. Bland–Altman method for measuring agreement

Predictor errors were evaluated with the Bland–Altman method. This method makes it possible to evaluate the agreement between two continuous variables described in detail in one of the most cited publications of all time [27]. The predictors area and median distance were measured based on a series of CT scans and based on only one CT slice in a dataset of 100 patients. This design allowed us to make a comparison between the two methods (predictor values based on series of CT scans and/vs only one CT scan) and 95% limits of agreements were calculated.

2.3.5. Kolmogorov–Smirnov confidence band

It was very important to investigate not just the difference in dose values, but the ‘absolute’ LAD dose values in the preferable position as well. For example, a 2 Gy LAD mean dose difference can be a result of 2 Gy and 4 Gy doses or a result of 20 Gy and 22 Gy doses. LAD mean dose value in the preferable position is the same as the minimum of LAD mean dose values in the prone vs. supine positions. Therapy plans in both positions were constructed, and knowledge of the LAD dose values in both positions made it possible to evaluate ‘absolute’ dose values as well. The empirical cumulative distribution function (EDF) of the ‘absolute’ dose values in the preferable position was investigated with the Kolmogorov–Smirnov method [28-30]. A 95% confidence band for the EDF was constructed, and dose constraints were determined based on percentile estimation.

3. Results

In this chapter, we present the steps of model-finding, starting with an examination of possible predictors, then results of the possible models together with model-checking, ending with the final model and its validation.

3.1. Results for the behaviour of the predictors

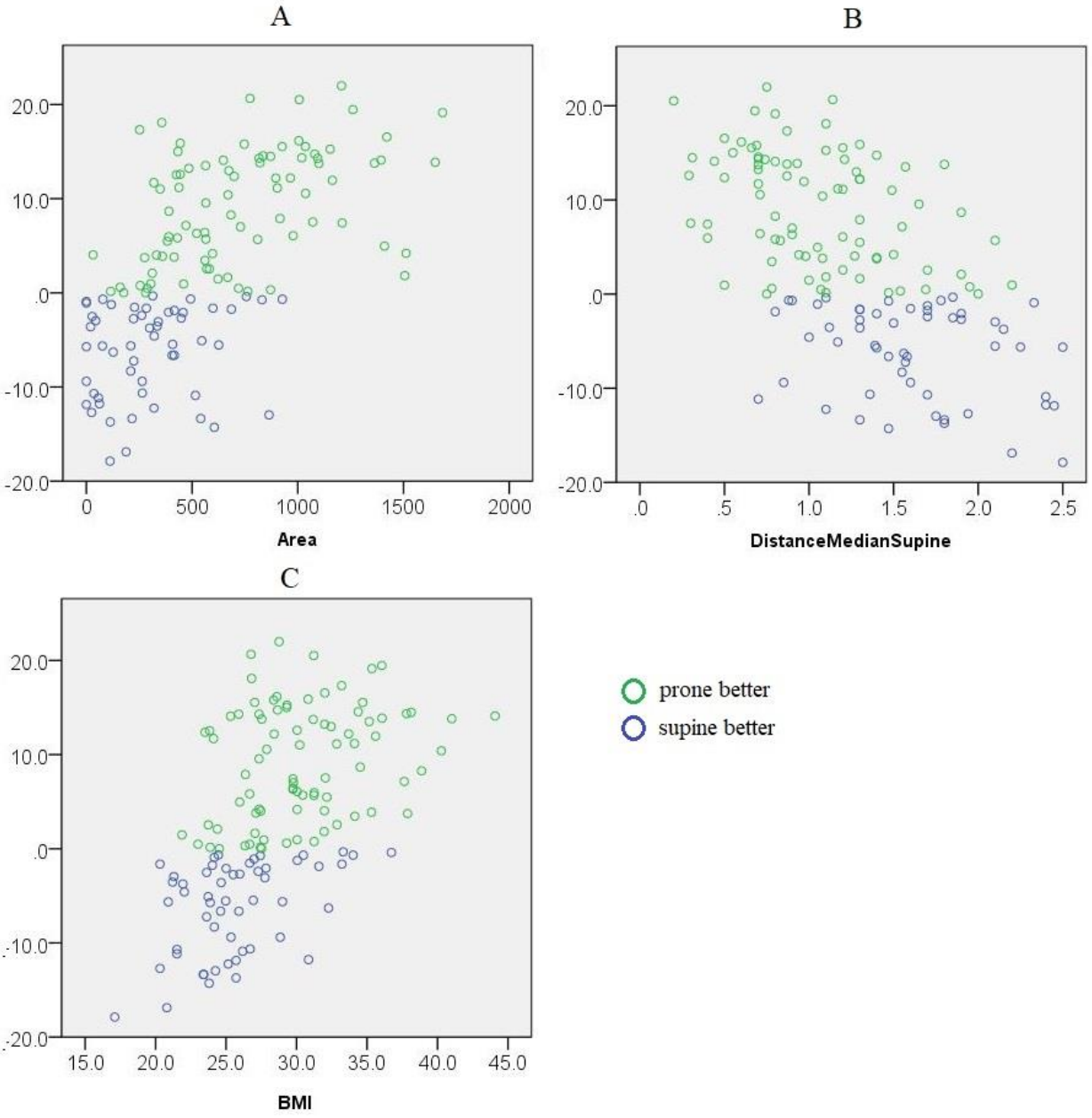


Figure 3. Scatter plots to present the relationship between the different predictors and the dependent variable. A refers to the predictor area, B shows the predictor median distance, and C illustrates the predictor BMI.

The first aim of this thesis is to investigate the possible relationship between the candidate predictors and the primary dependent variable. Scatter plots and ROC curves are presented.

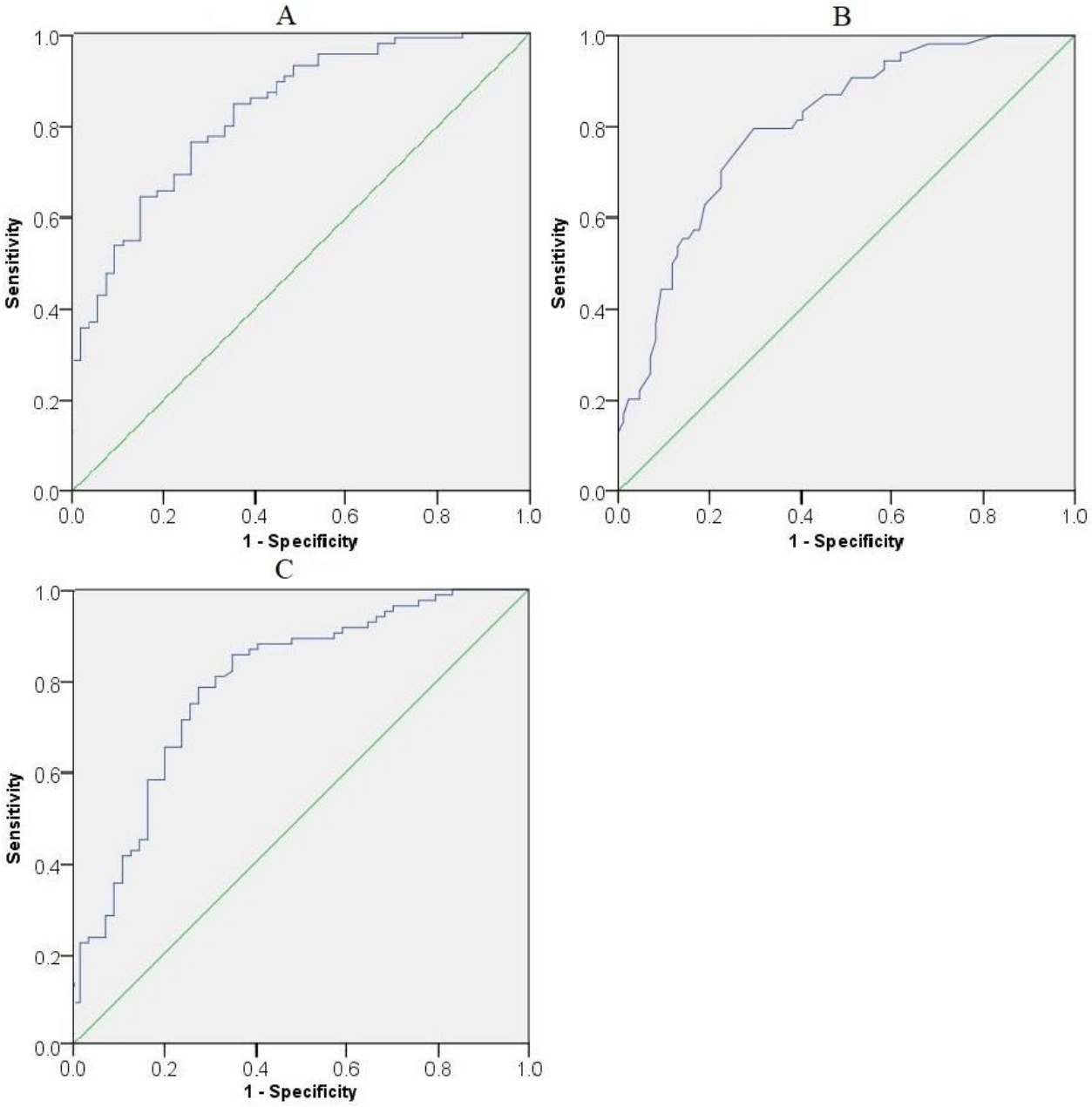


Figure 4. ROC curves for our predictors. A refers to the predictor area, B shows the predictor median distance, and C illustrates the predictor BMI.

Scatter plots present the relationship between the predictors and the dependent variable (LAD mean dose difference, ΔMD_{LAD}). It can easily be seen that none of the predictors alone seems to be sufficiently effective to make predictions.

These graphs show that it is quite challenging to make predictions based on these three predictors. There is a relatively large overlap between the points belonging to the groups.

The performance of the predictors alone was also evaluated with ROC analysis. The results are presented in Figure 4.

As was expected from the scatter plots, the ROC curves reflect a relatively good separation, which is not good enough for a prediction. The best predictor is area with AUC_{ROC}=0.868, 95% CI: 0.791, 0,944, while the predictor median distance is characterized as AUC_{ROC}=0.787, 95% CI: 0.690, 0.884. The 'worst' of these three predictors is BMI with AUC_{ROC}=0.740, 95% CI: 0.630, 0.850.

3.2. Prediction models

The second aim of this study is to develop a model-based classifier method to predict the preferable treatment position. Based on the scatter plots and ROC curves, it was obvious that none of the predictors alone can be used to make predictions, and it seemed quite challenging to predict a preferable treatment position based on these predictors.

First of all, we would like to avoid the use of strongly correlated predictors. Strongly correlated predictors in a multivariate model can lead to wrong parameter estimates (with large standard errors) and wrong conclusions [8, 10]. To avoid multicollinearity in a possible multivariate model, a hierarchical cluster analysis was performed with distance definition of the Pearson correlation coefficient. The correlations were explored among the predictors (Figure 5), and poorly correlated predictors were chosen for model building.

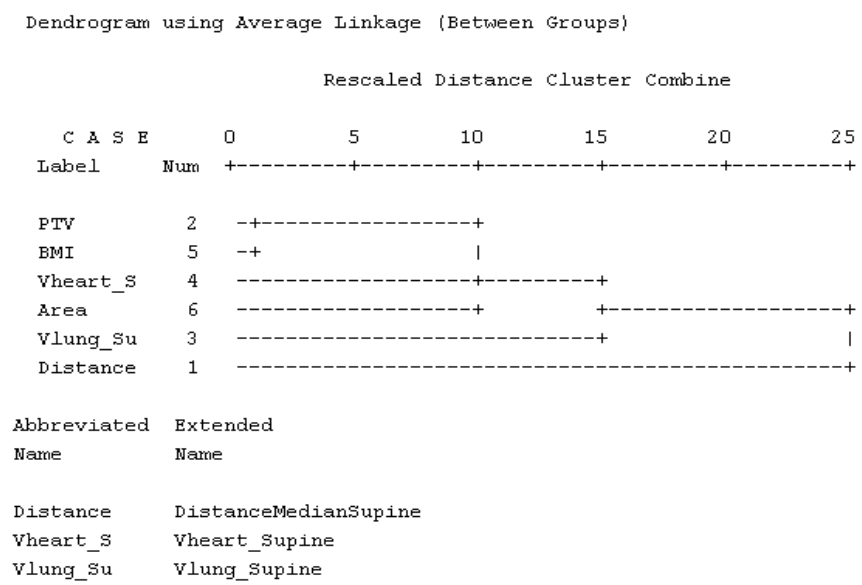


Figure 5. Result of a hierarchical cluster analysis (dendrogram) based on Pearson correlations.

It can be seen that PTV and BMI are strongly correlated ($r=0.681$, $p<0.001$) and BMI is easier to measure, so the use of BMI instead of PTV seemed to be reasonable [24].

3.2.1. Backward likelihood ratio selection model

Our early models were logistic regression-based models using patients' characteristics. The continuous dependent variable had to be dichotomized for the logistic regression-based models. We coded it 0 if the prone position was preferred and 1 if the supine position was preferred, based on the gold standard method (i.e. CT series and therapy plans in both positions).

Backward and forward likelihood ratio logistic regressions were used, based on the remaining, relatively uncorrelated variables. The backward likelihood ratio selection model resulted in the 'main effect model', as it contained the predictors without interaction terms:

area + BMI + median distance.

Estimated parameters of the model

The predictor area was originally measured in mm^2, but we changed it to cm^2 for a similar magnitude of parameter estimates.

	Estimated coefficient	S.E.	p
DistanceMedianSupine	-1.875	0.575	0.001
BMI	0.238	0.067	<0.001
Area_cm2	0.353	0.097	<0.001
Constant	-5.357	2.001	0.007

Table 1. Estimated coefficients of the backward likelihood ratio selection logistic regression model.

All parameters are significant at the 1% level. Although standard errors of the parameter estimates are rather large, the fitted model is stable and suitable for predictions.

ROC analysis for the predicted probabilities

It is possible to apply ROC analysis to the predicted values (predicted probabilities) of the logistic regression model (main effect model). It can be seen in Figure 6 that the ROC curve is promisingly close to the ideal curve. AUC_{ROC}: 0.906, 95% confidence interval for AUC_{ROC}: 0.854, 0.959.

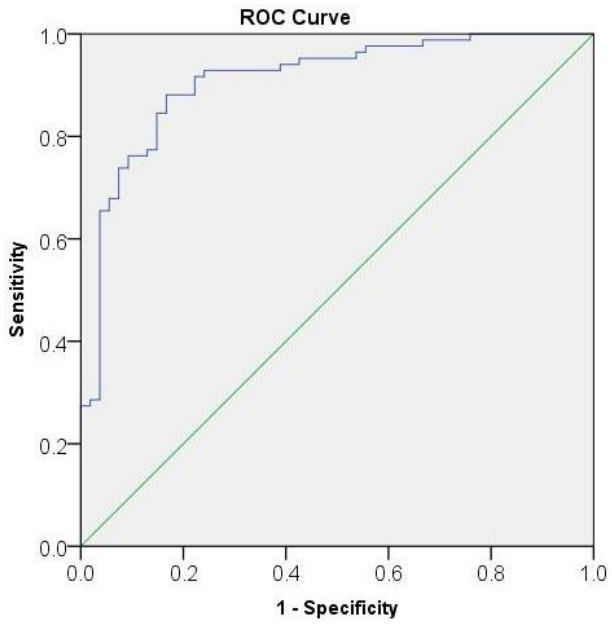


Figure 6. ROC analysis results for the logistic regression predicted probabilities. Hierarchical logistic model.

3.2.2. Forward likelihood ratio selection model

A very plausible method for model building is the forward likelihood ratio (LR) model selection.

The forward likelihood ratio selection model was not a hierarchical model, as it contained two interaction terms, without the ‘single’ terms:

area*BMI + area*median distance

without main effects.

Although non-hierarchical models can be difficult to interpret because they can contain interaction terms without the main effects, our goal is rather practical use than a deep interpretation of the coefficients.

Estimated parameters and equation of the model

	Estimated coefficient	S.E.	p
Area by DistanceMedianSupine	-0.369	0.114	0.001
Area by BMI	0.033246	0.006764	<0.001
Constant	-1.653	0.440	<0.001

Table 2. Estimated coefficients of the forward likelihood ratio selection logistic regression model.

A summary of the parameter estimates of this model is shown in Table 2. All parameters are significant at the 1% level.

ROC analysis for the predicted probabilities

It can be seen in Figure 7 that the ROC curve is promisingly close to the ideal curve. AUC_{ROC}: 0.900 with standard error 0.027, 95% confidence interval for AUC_{ROC}: 0.848, 0.953.

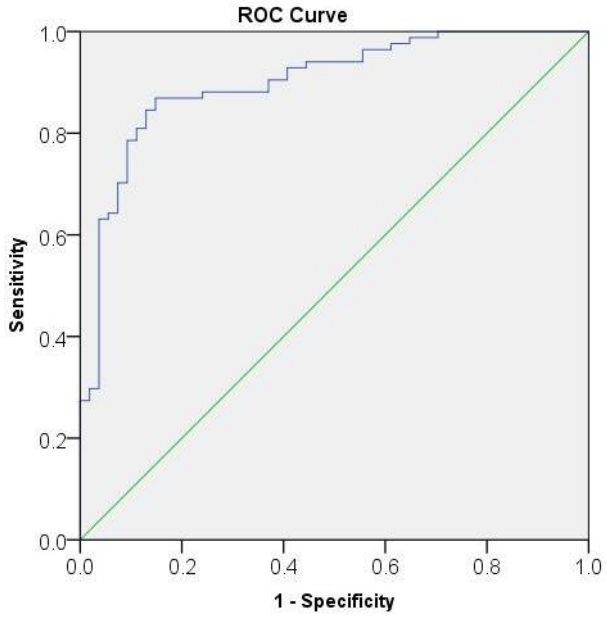


Figure 7. ROC analysis results for the logistic regression of predicted probabilities. Non-hierarchical model.

3.3. Linear regression-based model

A serious weakness of logistic regression-based classification methods is that the dependent variable is always binary. In our dataset, this would have led to a serious loss of information. The dependent variable, LAD mean dose difference, is a continuous variable which is dichotomized by its sign.

If we treat it as a continuous variable, the problem will be a ‘regression type’ problem. Higher differences will be taken into account with more weight. The value of the LAD mean dose

difference is also expected. Knowledge of the expected dose difference is informative for physicists and physicians involved in radiation treatment planning.

Estimated parameters of the model

Table 3 presents the estimated coefficients of the multiple linear regression.

	Estimated coefficient	S.E.	t-test statistic	p
(Constant)	-10.532	4.151	-2.537	0.012
DistanceMedianSupine	-5.726	1.165	-4.917	<0.001
BMI	0.573	0.123	4.660	<0.001
Area_cm2	0.830	0.158	5.243	<0.001

Table 3. Estimated coefficients of the linear regression model.

Although standard errors of the parameter estimates were rather large, the fitted model was useful for predictions. Almost all the coefficients of the fitted linear regression model are significant at the 0.001 level (except the constant, which is significant at the 0.05 level).

Goodness of fit of the linear regression model

The goodness of fit of the model was examined with the following measures: the multivariate correlation coefficient $R=0.754$ and the square of the adjusted multivariate correlation coefficient $R^2_{adj}=0.560$.

ROC analysis for the linear regression model

It is also possible to apply ROC analysis to the linear regression model as well. We present the result of the ROC analysis for the predicted values of the multiple linear regression model. Like earlier logistic models, good results can be seen in Figure 8, where the ROC curve is promisingly close to the ideal curve. AUC_{ROC} 0.903, 95% confidence interval for AUC_{ROC}: 0.850, 0.957.

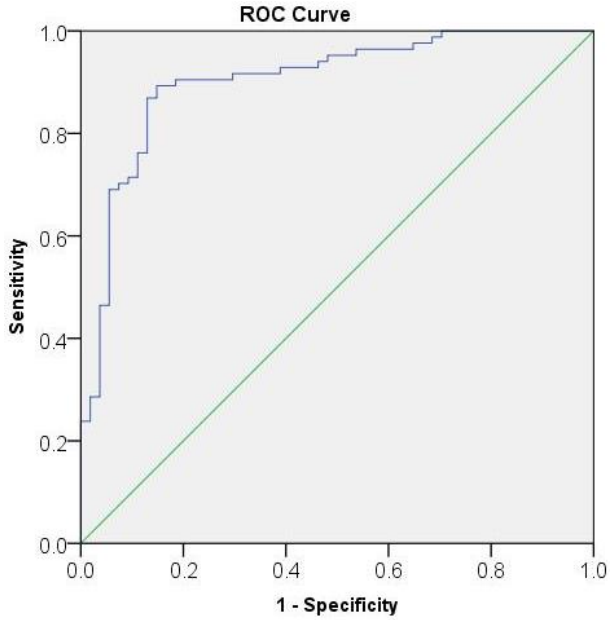


Figure 8. ROC curve for the linear regression-based predicted values.

This result clearly shows that the regression-based method is useful for prediction.

A summary of the results of these methods is shown in Table 4.

Model	AUC	95% Confidence interval for AUC	Brier score
BMI (predictor only)	0.740	0.630, 0.850	0.201
Area (predictor only)	0.868	0.791, 0.944	0.151
Median distance (predictor only)	0.787	0.690, 0.884	0.189
Logistic regression (main effect model)	0.906	0.854, 0.959	0.124
Logistic regression (forward LR selection model)	0.900	0.848, 0.953	0.132
Linear regression	0.903	0.850, 0.957	0.139

Table 4. Primary classification results for the prediction models.

3.3.1. Model diagnostics for the linear regression model

Despite the promising results in the previous sections, it is crucial to investigate the multivariate regression model in depth, especially to investigate the assumptions of the linear regression model [3].

Distribution of the residuals

The normality of the residuals was checked graphically with a Q–Q plot and a histogram with an estimated normal curve, which can be seen in Figure 9.

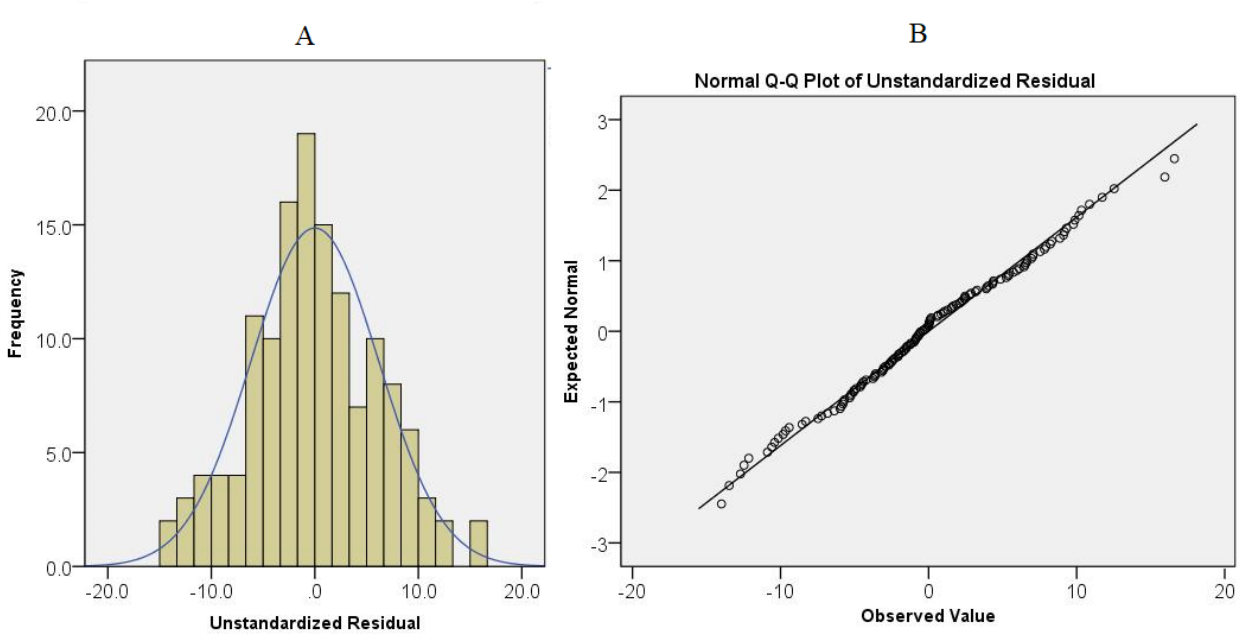


Figure 9. Distribution of the residuals: A. Histogram for the residuals with an estimated Gaussian (normal) density curve; B. Q–Q plot for the residuals.

The Q–Q plot reveals no large deviations and no tendency in deviations. The linearity of the points suggests that the residuals are normally distributed.

The Shapiro–Wilk test for normality confirms the result. The Shapiro–Wilk test results in a p-value of $p=0.592$, which confirms that we have no reason to assume a different distribution than normal.

Tendency of the residuals

The linear regression model assumes that the variance of the residuals is constant (see section 2.1.1.) The residuals might have an increasing (decreasing) or other kind of tendency in function of the predicted values. The plot in Figure 10 clearly demonstrates that there is no overall trend for the residuals, so the assumption of constant variance is fulfilled.

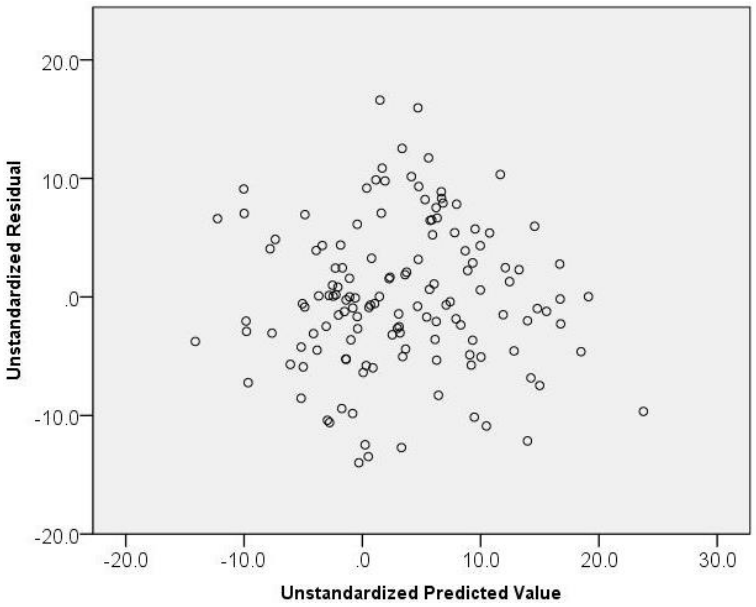


Figure 10. Scatter plot to present possible tendency of the residuals in function of the predicted value. The horizontal axis shows predicted values, while the vertical axis represents residuals.

The plot for the residuals in the sequence of the data collected and the result is presented below (Figure 11).

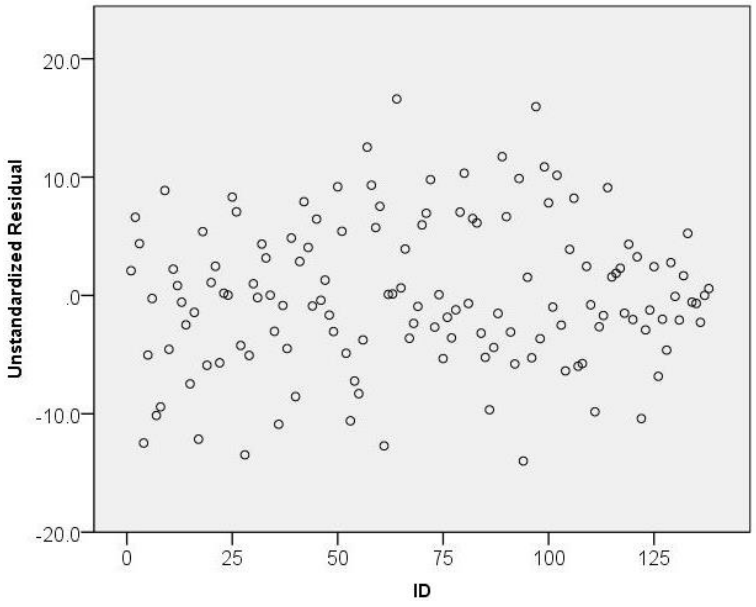


Figure 11. Scatter plot to present possible time-dependent tendency of the residuals. The horizontal axis shows the ID numbers of the patients, while the vertical axis represents residuals.

It can be clearly seen that there is no overall time-dependent tendency at all.

Independence of the residuals

The Durbin–Watson statistic based on a sample of 138 patients is *d=1.847.* 5% significance points for the Durbin–Watson test are $d_u=1.774$ $d_L=1.693$ if *n=150.* The acceptance interval is $d_u=1.774$ to $4-d_u=2.226$, so our decision is to accept the independence of the residuals [31, 32].

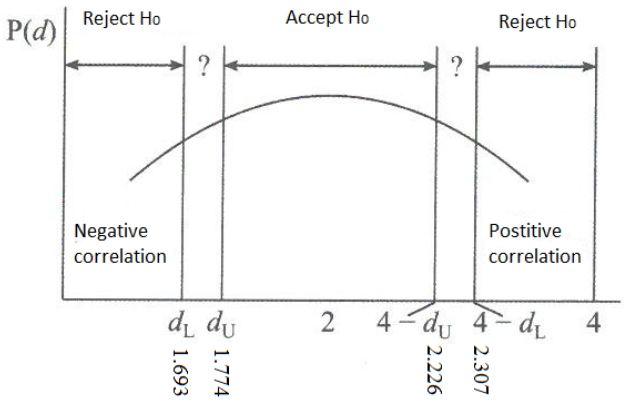


Figure 12. Acceptance and rejection regions for the Durbin–Watson test. This figure is a modified version of the original Figure 8-4 on page 202 of the reference [3].

3.3.2. Model validation

Multiple linear regression seemed to be the most useful model, since it provides an expectation of the dose difference. However, it is well known that the results of the classifications are overoptimistic when the classification is made on the same dataset where the classification rule was discovered (i.e. the training set and the test set are the same). We used an internal validation method of data splitting as follows: the sample was divided into two parts randomly, with 70% of the sample as the training set for the linear regression and the resultant model being tested on the remaining 30% of the data. The classification results and the misclassified dose were also noted. The process was repeated 1000 times randomly, and it was not just the proportions of misclassified patients that were taken into account, but also the distribution of the misclassified dose.

We found that a cut-off value of 0.6 was an optimal choice with respect to the expected misclassified dose values and high values as sensitivity of 80.7% and specificity of 87.5%.

Cut-off	Sensitivity (%)	Specificity (%)	Extent of wrong estimate, decision: prone (Gy, mean±Sd)	Extent of wrong estimate, decision: supine (Gy, mean±Sd)
-0.6	66.6	91.1	2.5±3.9	-0.7±1.0
-0.3	70.8	90.7	2.6±3.6	-0.8±1.1
0	74.4	90.0	2.4±3.4	-0.9±1.3
0.3	77.7	88.9	2.1±3.0	-1.2±1.6
0.6	80.7	87.5	1.7±2.6	-1.7±1.9
0.9	83.4	86.0	1.5±2.4	-2.0±2.2
1.2	85.4	83.6	1.1±2.3	-2.3±2.8
1.5	86.5	81.7	1.1±2.2	-3.0±3.7
1.8	86.8	79.9	1.3±2.3	-3.5±4.2

Table 5. Classification measures based on 1000 times random cross-validation. (We define supine position as positive)

3.4. A few results on decision curves for 'non-probability outcomes'

Our goal is to apply the decision curve method to linear regression-based predictions as well as logistic regression-based predictions and compare our models with respect to net benefit values. In the case of logistic regression models, the prediction is an expected probability; therefore, the DCA can be applied directly. Linear regression results in an expected dose difference, so the predicted value must be transformed into a range of probabilities.

The purpose of this section is to investigate how the shape of the decision curve depends on the transformation function chosen and the probability distribution of this non-probability outcome. These are unpublished results.

Although it is reasonable to use logistic regression to predict probabilities, there are other plausible methods to solve this problem. In other words, there are several possible link functions to transform the predicted scores or the expected values into probabilities. Plausible methods are, for example, the use of the empirical cumulative distribution function (CDF) to obtain probabilities or use an inverse logit link, probit link function or logistic regression-based probabilities.

These four transformations were chosen and examined for their effect on the shape of the decision curve. Simulations were performed for the four transformations applied to symmetrical and skewed probability distributions. The performance of the prediction naturally depends on the shift of the groups, i.e. on the shift parameter between the two groups.

A very fundamental doubt is that the shape of the decision curve depends on the transformation chosen. In this section, it will be confirmed that the shape of the decision curve depends on the transformation (link function) chosen.

3.4.1. Simulation 1: The shape of the decision curve depends on the transformation function chosen. Different continuous distributions

The first approach is the 'regression type approach'. 10 000 random samples were simulated. Random predictor (X) was simulated from different distributions: standard normal, uniform, gamma and lognormal with the parameters noted below in formulae (14)...(17). We added random residuals normally distributed with the parameters below. To construct a binary dependent variable, a continuous dependent variable (Y) was first calculated as the sum of the predictor variable (X) and the error term (ε).

10 000 random samples were simulated from normal (X_1), uniform (X_2), gamma (X_3) and lognormal (X_4) distributions. The added random residuals (ε) were normally distributed with mean=0 and SD=0.5. The four types of continuous dependent variables here are in (14), (15), (16) and (17).

$$Y_1 = X_1 + \varepsilon, \quad X_1 \sim N(0,1) \quad \text{and} \quad \varepsilon \sim N(0,0.5^2) \qquad (14)$$

$$Y_2 = X_2 + \varepsilon, \quad X_2 \sim Uniform(-1,1) \quad \text{and} \quad \varepsilon \sim N(0,0.5^2) \qquad (15)$$

$$Y_3 = X_3 + \varepsilon, \quad X_3 \sim Lognormal(1,1) \quad \text{and} \quad \varepsilon \sim N(0,0.5^2) \qquad (16)$$

$$Y_4 = X_4 + \varepsilon, \quad X_4 \sim Gamma(2,1) \quad \text{and} \quad \varepsilon \sim N(0,0.5^2) \qquad (17)$$

where X has standard normal, uniform, gamma and lognormal distributions, respectively, ε is the error term, ε has $N(0,0.5^2)$ distribution, and X and ε are independent.

In the second step, two groups were formed to obtain binary dependent variables.

The continuous dependent variables ($Y_1 \ldots Y_4$) were cut at their median values to achieve a binary response. For example, if the current value of Y_1 exceeded the median value, then this case was defined as positive; otherwise, it was defined as negative, and the 'prevalence' was therefore 0.5. Finally, the random predictors ($X_1 ... X_4$) were transformed into 'probabilities' using probabilities predicted with empirical CDF (model 1), inverse logit link (model 2), probit function (model 3) and logistic regression (model 4), and we constructed decision curves. The dependence of the binary response on the random predictors ($X_1 ... X_4$) was also examined with ROC curves.

Results are shown in Figure 13. The area under the ROC curve shows a quite good separation of the original predictors.

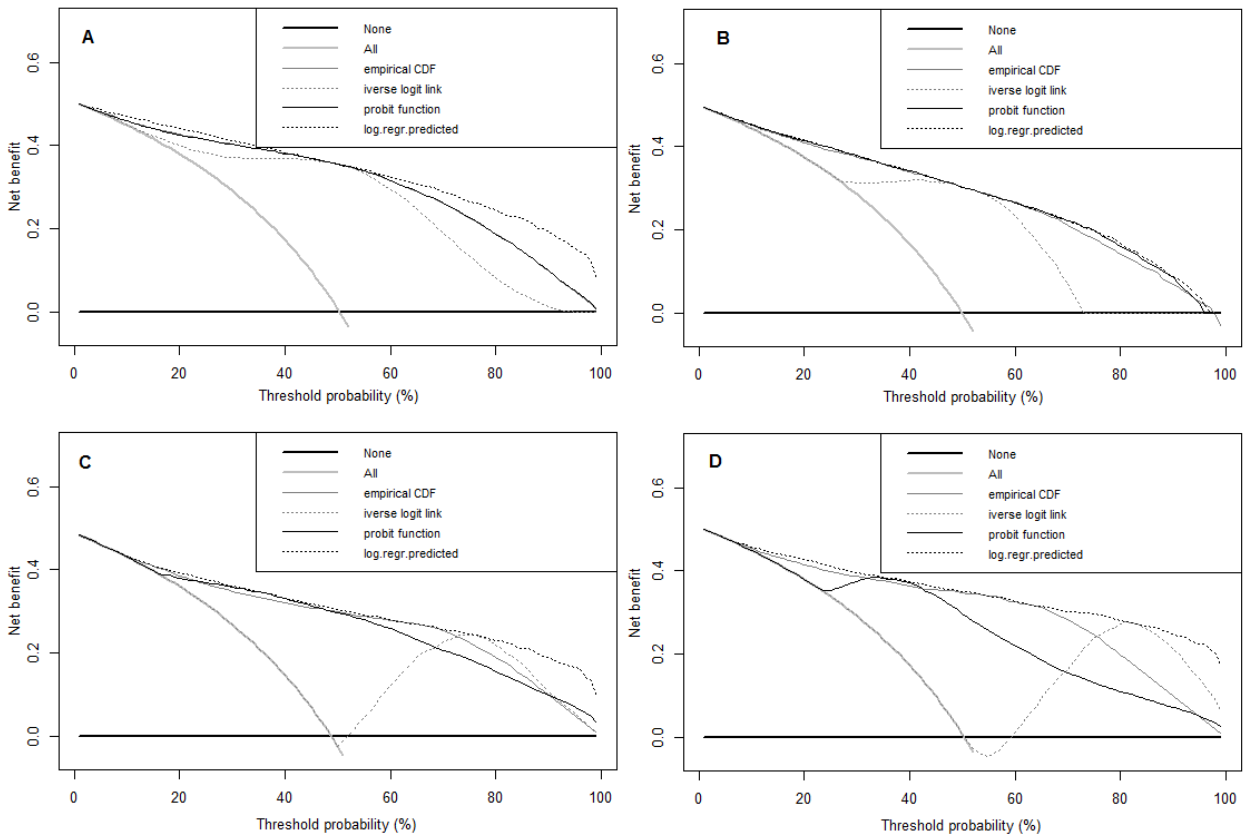


Figure 13. Simulation results using empirical CDF (model 1), inverse logit link (model 2), probit function (model 3) and logistic regression-predicted probabilities (model 4) transformation methods. A. The simulated score has a standard normal distribution. Skewness: 0.07, AUC: 0.947 (ROC). B. The simulated score has a uniform distribution in [-1,1]. Skewness: 0.03, AUC: 0.882 (ROC). C. The simulated score has a gamma distribution (shape=2, mean=1). Skewness: 1.715, AUC=0.95 (ROC). D. The simulated score has a lognormal distribution. Lognormal (1,1). Skewness: 2.948, AUC: 0.925 (ROC).

Figure 13A–B refers to two symmetrical distributions. The curves are quite close to each other for a standard, normally distributed score. In Figure 13B, it is clear, that only model 2 (inverse logit link) differs from the others. Figure 13C–D illustrates two specific skewed distributions, the gamma distribution and lognormal distribution. We conclude that the different transformation functions diverge for these skewed distributions.

Not surprisingly, model 2 (i.e. probabilities predicted by the inverse logit link) led to the most pessimistic and non-monotonic results for skewed distributions (this is especially noticeable in Figure 13C–D). Model 4 (logistic regression predicted probabilities) produced the most optimistic values of net benefit in all cases.

3.4.2. Simulation 2. The shape of the decision curve depends on the transformation function chosen. Two normal distributions.

In this section, the simulated score has a standard normal distribution for group 0 (the non-diseased or negative group) and different normal distributions (shifted) for group 1 (diseased or positive). With this approach, we assume that the diseased group also has a normally distributed score value, but with a different mean than the non-diseased group. The prevalence is 0.2. In this section, the performance of the classification depends on the shift parameter between the normal distributions.

10 000 random samples were simulated. The behaviour of the four transformation functions was then examined. We conclude that empirical CDF (model 1), inverse logit link (model 2) and probit function (model 3) lead to extremely low values of net benefit despite the very good classification. This is especially noticeable in Figure 14D, where the shift parameter is 3SD between the normal distributions, which represents a very good classification of the cases. Logistic regression predicted probabilities produced the best values of net benefit as in section (3.4.1). Results are shown in Figures 14A–D.

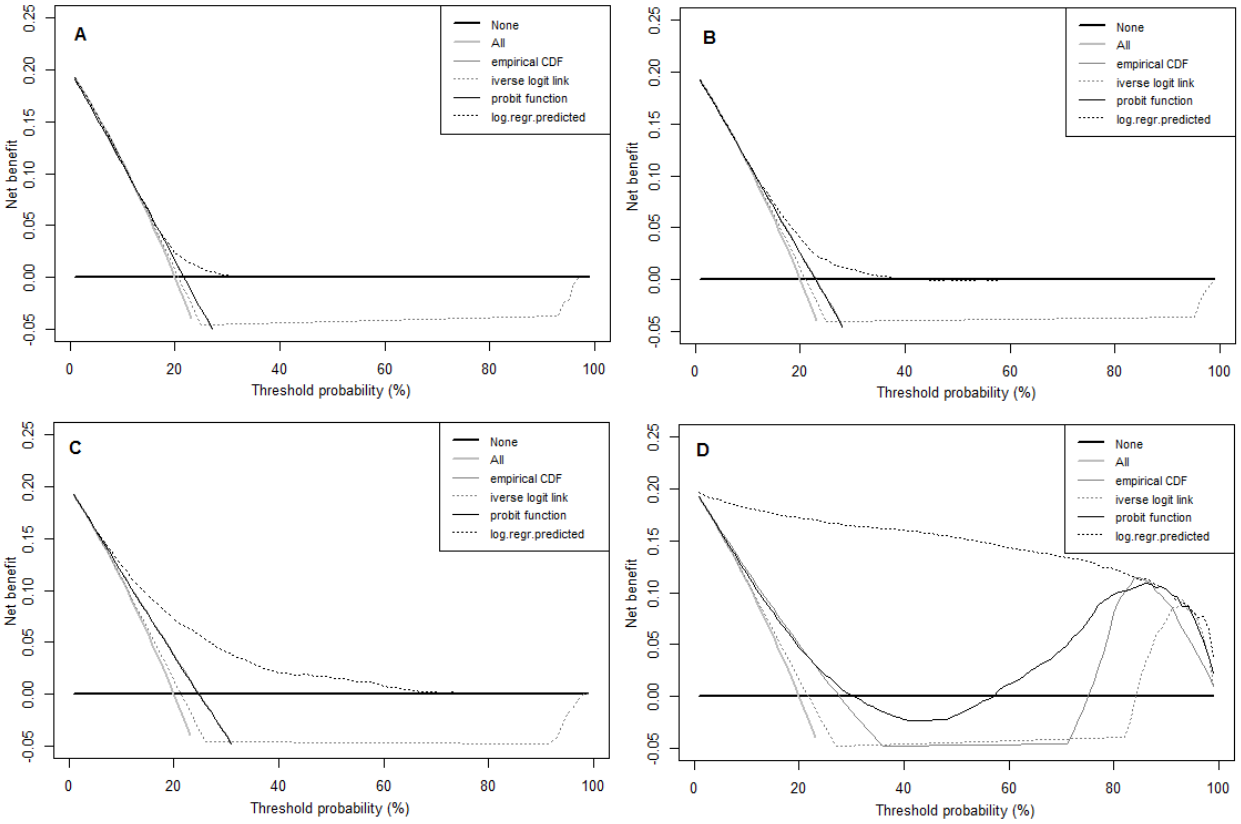


Figure 14. Simulation results, two normal distributions. Empirical CDF (model 1), inverse logit link (model 2), probit function (model 3) and logistic regression predicted probabilities (model 4). A. The simulated score has a standard normal distribution in the non-diseased group and a normal distribution with mean=0.33 and SD=1 in the diseased group, AUC: 592 (ROC). B. The simulated score has a standard normal distribution in the non-diseased group and a normal distribution with mean=0.5 and SD=1 in the diseased group, AUC: 0.639 (ROC). C. The simulated score has a standard normal distribution in the non-diseased group and a normal distribution with mean=1 and SD=1 in the diseased group, AUC: 0.762 (ROC). D. The simulated score has a standard normal distribution in the non-diseased group and a normal distribution with mean=3 and SD=1 in the diseased group, AUC: 0.983 (ROC).

It can be concluded based on these results that models 1–3, i.e. empirical CDF (model 1), inverse logit link (model 2) and probit function (model 3), can lead to very low values of net benefit even if the classification is very good and that these three transformation functions can bring about very different results, especially for skewed distributions.

3.4.3. Conclusion about non-probability outcomes

Based on both approaches, we recommend the use of logistic regression-based probabilities to construct the decision curve for 'non-probability outcomes'. Our recommendation is the same as in the R package by Vickers and Elkin [2]. We confirm that the most appropriate transformation for such cases is logistic regression, which is used in the DecisionCurve and rmda R packages [2, 33].

The decision curve method is applicable to continuous outcomes as well. There are two approaches to this question.

- Decision curves could be regarded as only applicable to probability expectations, and the link function is part of the model that needs to be evaluated.
- The link function could be regarded as part of the evaluation method, and the outcome of the model is a score or continuous expected value.

These are two different aspects, but the conclusion is the same. We also recommend the use of logistic regression-based probabilities to construct the decision curve for 'non-probability outcomes'.

3.5. Comparing our models using decision curves

The simulation results noted above allow us to evaluate the linear regression-based model with respect to net benefit and DCA as well. The simulation results confirm the use of logistic regression-predicted probabilities to construct decision curves for 'non-probability outcomes'. We examined the performance of the three best predictors alone with decision curves, and we compared them to the models described in (3.2) and (3.3). None of the best three predictors alone was comparable to the prediction models (Figure 15). For example, a univariate logistic regression was applied with the independent variable BMI (and with the dependent variable preferable treatment position, we defined supine as positive) to construct probability prediction based on the value of BMI. In Figure 15 'all' refers to 'all positive' (i.e. all patients treated in the supine position) and 'none' indicates 'all negative' (i.e. all patients treated in the prone position).

In our analyses, decision curves for both the logistic regression model and the linear regression-based model were quite similar to each other; additionally, it is important that both models led to high values of net benefit for a wide range of threshold probabilities. In other words, it is beneficial to use these models in respect of net benefit regardless of the current threshold probability. These results showed that these models can be used in clinical practice.

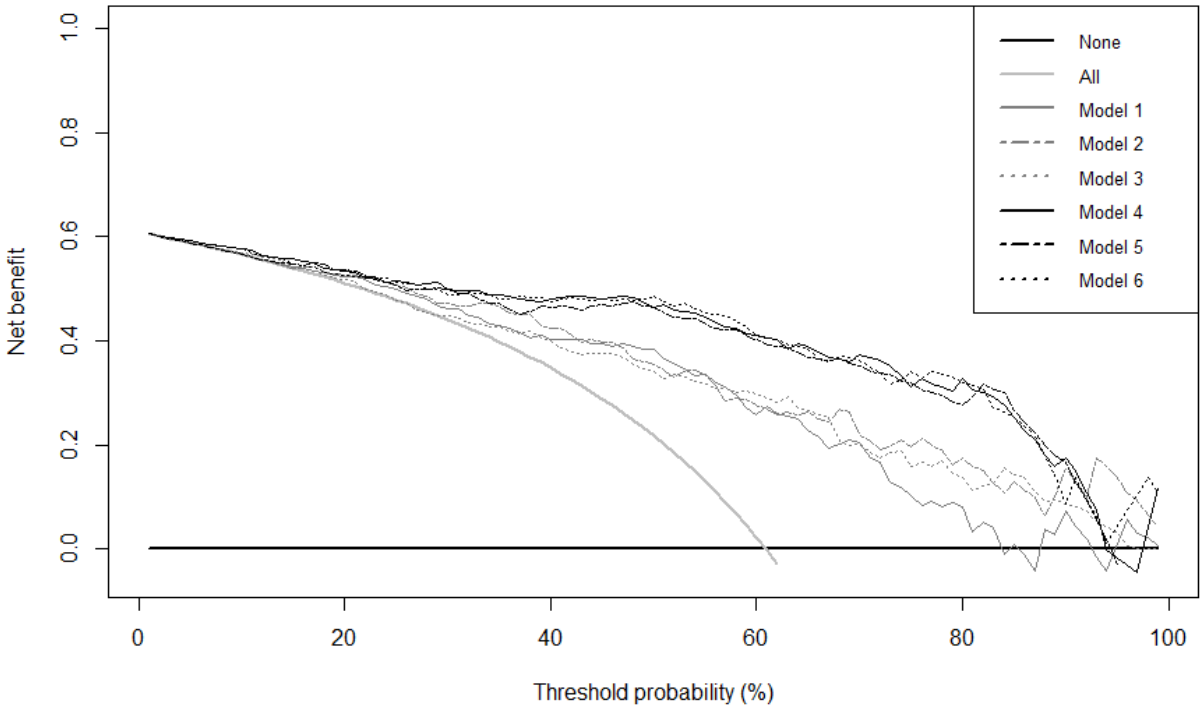


*Figure 15. Decision curves for the three best predictors and three models. Model 1 refers to the predictor BMI, model 2 to the predictor area, and model 3 to the predictor median distance; model 4 and model 5 are logistic regression-based models. Model 4 is the main effect model (area+BMI+median distance), model 5 is the forward likelihood ratio selection model, with two interaction terms (area*BMI+area*median distance), and model 6 is the linear regression-based model.*

A bootstrap method was applied to construct 95% confidence intervals for the net benefit at various threshold levels for the main effect model. 10 000 bootstrap samples of sample size 138 were generated by Microsoft Excel. Net benefit values were calculated by definition. The 2.5% and 97.5% percentiles of the net benefit values were calculated at each threshold level. The results are shown in Table 6.

Threshold probability	Lower bound for net benefit	Upper bound for net benefit
0.1	0.564	0.573
0.2	0.496	0.547
0.3	0.450	0.527
0.4	0.423	0.517
0.5	0.391	0.507
0.6	0.319	0.467
0.7	0.256	0.435
0.8	0.210	0.435
0.9	0.000	0.333

Table 6. 95% confidence intervals for net benefit (logistic regression, main effect model) at different threshold levels. Confidence bounds are based on 1000 bootstrap samples.

The differences between the logistic regression models and the multiple regression-based model are not relevant with respect to the very similar net benefit values. All of these models led to high values of net benefit for a wide range of possible threshold levels.

3.6. Predictor error consequences and further testing on an independent dataset of patients

In all of the previous sections, we assumed that there are no predictor errors at all. This was quite plausible because a series of CT scans can provide a precise way to measure the predictor values. In this section, we introduce the predictor values based on only one CT slice. It is very reasonable to assume that predictor values based on only one CT slice might be affected by measurement errors.

In this section, we would like to reach two of our goals. First is the investigation of possible predictor errors, and second is the testing of the linear regression-based model in the data for 100 more patients. We found it very important to examine the possible predictor measurement errors. The selection of the single CT slice might have an effect on predictor values. This phenomenon can be called a ‘plane miss’.

The author of this thesis is not an expert in CT methodology and treatment planning methods. The selection method of the single CT slice and the ‘plane miss’ phenomenon is described in detail in [26, 34].

At this stage (we can call it stage 3, see the description in the data section (2.3.1)), our linear regression-based model was intended to use the predictor values area and median distance based on only one CT slice.

If this model works well, a ‘calculator’ can be prepared to aid physicians in decision-making in their practice [26].

Before everyday use of the calculator, it is essentially important to investigate these errors and especially investigate their consequences in the prediction [26].

3.6.1. Predictor errors

The predictor errors for area and median distance are evaluated here. These predictors were measured on the CT series (as a reference) and based on only one CT slice in the sample of 100 patients. This provides the opportunity to compare the values.

Bland–Altman plots were used for comparison; they can be found in Figure 16. 95% limits of agreement are -5.624 cm^2 to 3.902 cm^2 for the predictor area and -0.829 cm to 0.783 cm for the predictor median distance.

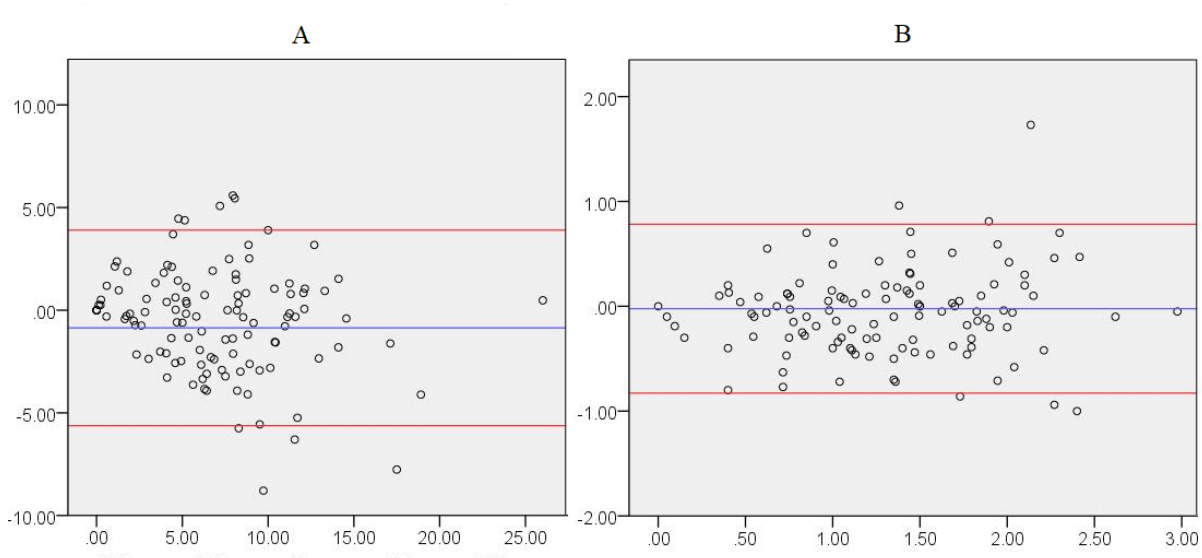


Figure 16. Bland–Altman plots for predictor error evaluation. A refers to the predictor area, and B illustrates the predictor median distance. The horizontal axis represents the average for the two different measures (i.e. the CT series vs. one CT slice), while the vertical axis shows the difference between the two measures. The red lines refer to 95% limits of agreement.

Although these predictor errors are reasonably high, we will see that the effect on the predictions is not critical.

3.6.2. Effect of predictor errors in the final prediction

First, the Bland–Altman plot for the predicted values will be presented. Here we calculated the predicted values with our ‘calculator’ (i.e. the linear regression-based model) based on the predictor values from only one CT slice and based on the predictor values from the CT series.

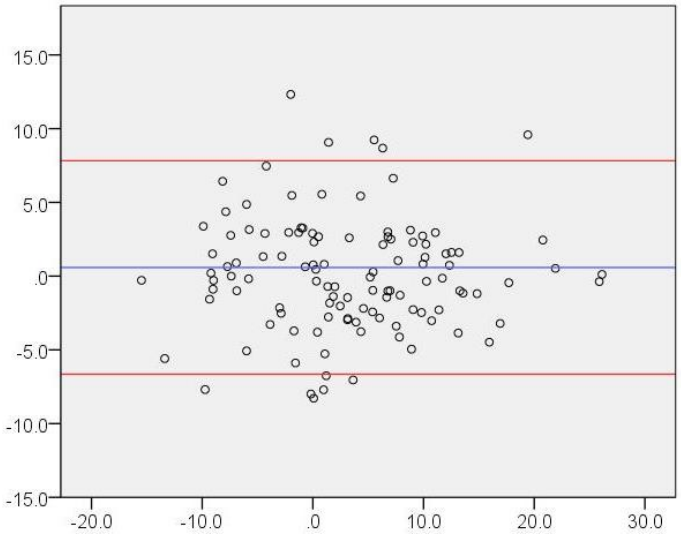


Figure 17. Bland–Altman plots for prediction error evaluation. This figure compares the predictions of the linear regression-based model (i.e. ‘calculator’) based on only one CT slice and a series of CT slices. The horizontal axis represents the average of the predictions, while the vertical axis shows the difference between the two predictions. Red lines refer to 95% limits of agreement.

95% limits of agreement are -6.65 Gy to 7.82 Gy. These limits define quite a long interval. The Bland–Altman plot seems to be the most important graphic method to show agreement between two measures [27, 35]. However, here our goal is prediction, and Figure 18 below is at least as important as the Bland–Altman plot was. Figure 18A shows a comparison of predictions based on a series of CT slices and based on only one CT slice, and this is equivalent to Figure 17 above, but in Figure 18B the horizontal axis refers to the prediction based on only one CT slice and the vertical axis presents the DVH data (from radiation treatment plans). DVH data is regarded as the gold standard to ascertain the real LAD dose difference.

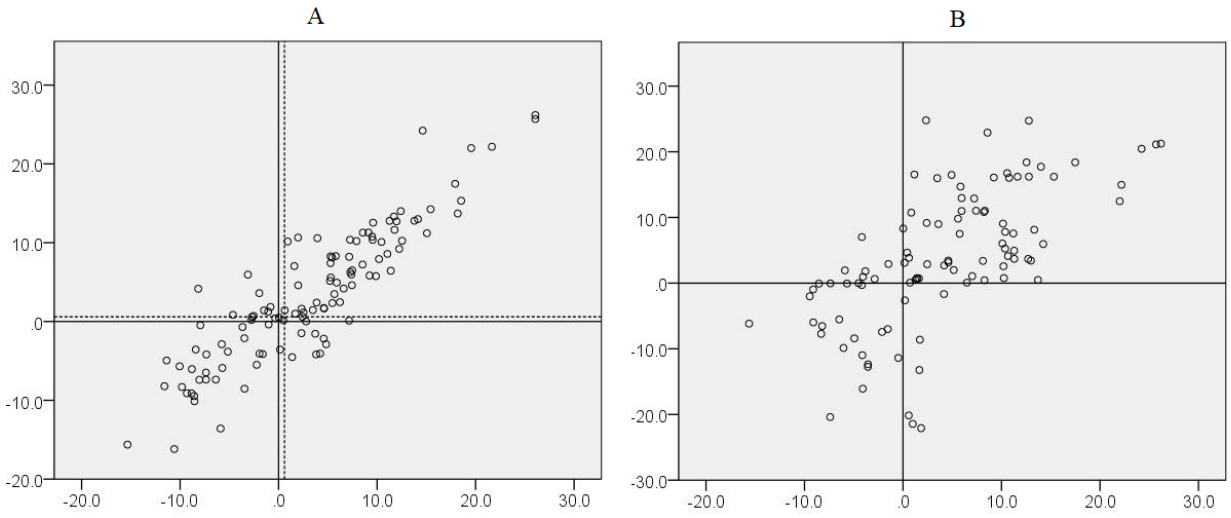


Figure 18. Comparison of predictions. A. Linear regression-based predictions based on a single CT scan (horizontal axis) vs. linear regression-based predictions based on a series of CT scans (vertical axis). B. Doses according to the estimation of the simple clinical method (linear regression) based on a single CT scan (horizontal axis) vs. DVH data (gold standard) extracted from the planning system (vertical axis). Dashed lines indicate cut-off values of 0.6 Gy. [26]

Although the Bland–Altman limits of agreement were reasonably wide, Figure 18B is promising and most of the points are located in the 1st and 3rd quadrants, suggesting that a large portion of the predictions will be correct as well. Table 7 contains the validation data, and it is important to compare the second and third columns of this table to the last two. This table clearly demonstrates that the effect of possible predictor measurement errors is not crucial in our simple predictive tool [9].

	Original patient data (double CT method) cross-validation results		Using predictors based on only one CT slice	
Cut-off	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
-0.6	66.6	91.1	72.4	91.5
-0.3	70.8	90.7	75.9	91.5
0	74.4	90.0	75.9	91.5
0.3	77.7	88.9	79.3	88.7
0.6	80.7	87.5	82.8	87.3
0.9	83.4	86.0	82.8	83.1
1.2	85.4	83.6	86.2	81.7
1.5	86.5	81.7	86.2	77.5
1.8	86.8	79.9	93.1	76.1

Table 7. Classification measures for ΔMDLAD using a single discrimination threshold. Great consistency is seen between the original cohort and the present series [26].

3.6.3. Empirical distribution of LAD dose with confidence bands: Dose constraints

An important practical aspect of this breast irradiation positioning problem is that the LAD mean dose difference can be very small even if the actual LAD dose values are quite high. Therefore, empirical distribution of the LAD dose values is fundamental. Therapy plans in both positions were generated (gold standard method), and the knowledge of the LAD dose values in both positions made it possible to evaluate 'absolute' dose values as well. The dose in the preferable position is the minimum of the dose planned values in the prone and supine positions.

Practical dose constraints can be defined based on this empirical distribution. 95% confidence bands were constructed with the Kolmogorov–Smirnov method [28-30].

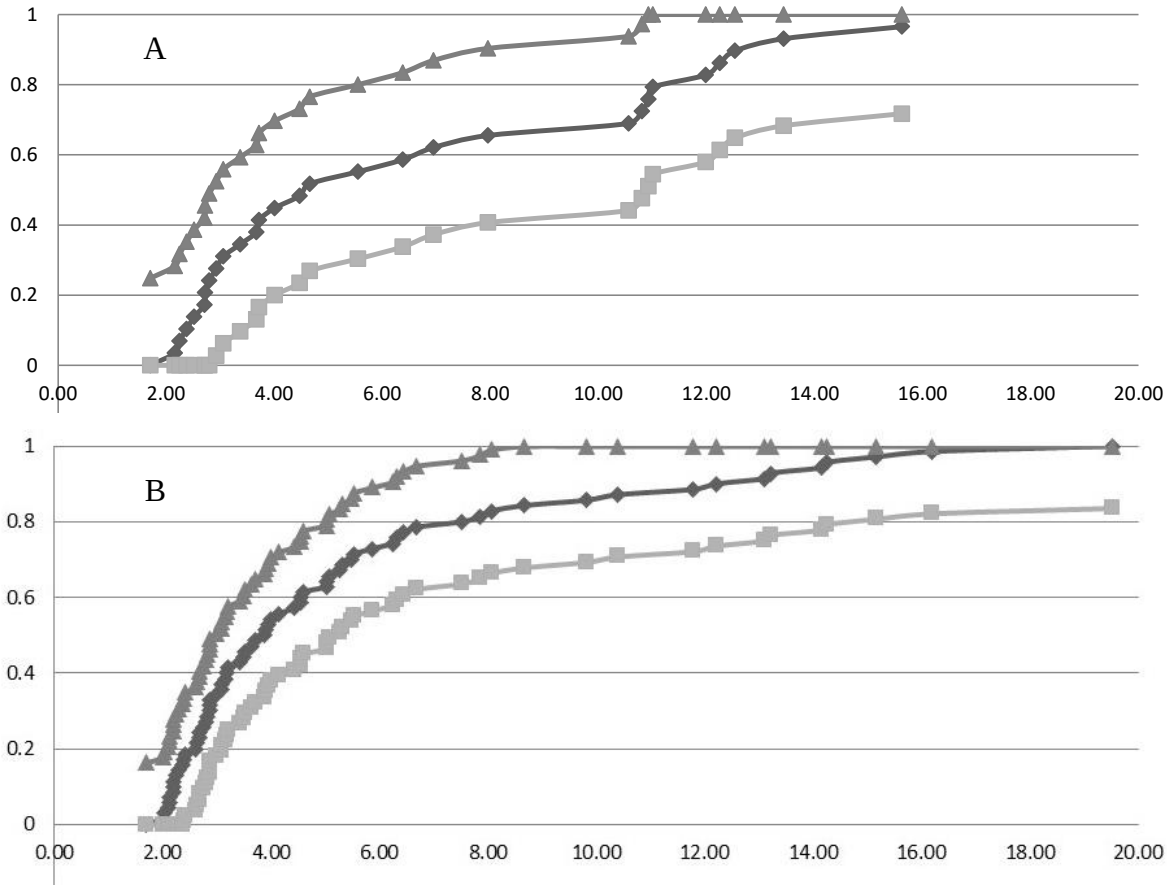


Figure 19. Empirical cumulative distribution function (EDF) for LAD dose values in the preferable treatment position with 95% confidence bounds. Figure A refers to the supine position and B to the prone position. The horizontal axis indicates the LAD dose value in the preferred position, while the vertical axis shows the empirical probability. This figure is based on the 100 patients' data in the stage 3 validation set.

Dose constraints were defined based on the empirical distribution and based on medical considerations. We agreed on the use of the 90% percentile of the EDF of the dose values. Our estimates were 12.5 Gy and 12.9 Gy in the supine and prone positions, respectively [26].

Although Kolmogorov–Smirnov confidence bounds are reasonably wide, in the supine position even the lower bound of the 95% confidence band is close to 70% at the point of 12.5 Gy. This means that it is very unlikely that more than 30% of the cases in the future will exceed this value. In prone position the lower bound of the 95% confidence band is also around 70% at 12.9 Gy, so the conclusion is very similar; it is also very unlikely that more than 30% of the patients will exceed this 12.9 Gy value in the prone position.

We determined dose constraints based on the empirical distribution of the dose values in the preferable position in the last section. We regarded these dose constraints as reasonable bounds above which another kind of intervention is needed. This other kind of intervention may be treatment planning in the other position or application of other radiation therapy techniques, such as IMRT (intensity-modulated radiotherapy) [26].

3.6.4. External testing of the model

The validation process can be divided into three different steps [9]. The results of the random cross-validation can be regarded as ‘internal validation’ [24, 25]. The classification results for the additional 100 patients’ data and the results based on a single CT slice can be considered as a ‘temporal validation’ because the data was derived from the same centre at a later time. External validation is needed to present generalizability of the model. One successful step was the external testing of our linear regression-based model. It was tested in a 28-case dataset of left-side breast cancer patients from Liège and showed great consistency with our results noted above. The predicted treatment position was correct in 24/28 (accuracy: 85.7%) cases [26, 34].

4. Discussion

Therapeutic prediction models were developed and evaluated in our interdisciplinary study. These models were based on well-established medical statistical methods, such as linear regression and logistic regression [1, 36-39].

The mathematical aspects of the evaluation of the predictive power of these models were presented in detail. The novel method of decision curves was used to compare these models. Simulations were carried out to clarify how the DCA can be performed on 'non-probability outcomes', such as scores or expected dose values. An evaluation of the predictive model was presented in detail. The widely accepted method of cross-validation was used for internal validation. Temporal validation and investigation of the possible predictor errors were carried out using an additional dataset of patients. The oft-cited method developed by the two great Russian mathematicians, Kolmogorov and Smirnov [28-30], was applied to construct a confidence band for the empirical cumulative distribution of the dose values. This method was used to determine dose constraints.

None of the single predictors alone was satisfactory for prediction. All the predictive models performed better than single predictors did. The linear regression model was considered the most clinically relevant for quantitative estimation, since it provides an estimate for dose difference and is comparable to the two logistic regression models in all other aspects. DCA demonstrated very similar performance, and AUC_{ROC} was almost the same for both the linear regression model and the logistic regression-based models. Validation steps confirmed that the model is stable. The Kolmogorov–Smirnov method showed that it is very likely (i.e. there is a 95% chance) that less than 30% of patients exceed the dose constraints we defined.

Strengths and weaknesses of the study

It is very important to minimize OAR radiation doses with individual positioning. The linear regression-based tool estimates the difference in the expected dose values based on the BMI and the d_{median} and A_{heart} measured on a CT slice at the middle of the heart. The result was compared to that of the full CT series in both positions and the dosimetric data. The comparison between the linear regression-based tool (based on only one CT slice) and the original method (predictors based on series of CT-scans) found very consistent results (very similar sensitivity and specificity values) [26]. This linear regression-based model and the calculator can be used in everyday practice as a useful tool to aid in radiation treatment planning.

One strength of our study is its relatively large sample size and multivariate aspect. Decision curves and the AUC for the ROC results were found to be similar for the linear regression model and the two logistic regression models [25]. The logistic regression models weight the outcomes on a binary scale, while the linear regression model weights them in keeping with the magnitude of the difference, which we regarded as a great advantage for practical aspects. Since the linear regression model provides additional information, i.e. an estimate of the dose difference, we decided to use the linear regression model. Knowledge of the estimated quantitative benefit of one or the other treatment position during radiotherapy may provide better guidance for the physician when considering various aspects, such as repositioning accuracy and patient comfort.

Applying AUC_{ROC} and measures like sensitivity and specificity is very common in radiotherapy planning, but we have not seen the approach of using decision curves in this field. Our investigations point to the clinical utility of predictive models. The methodology described in this thesis might be adapted for decision-making problems in other areas of medicine.

This investigation has its own limitations. One limitation might be that we assumed a linear relationship between the predictors and the dependent variable. It can be seen in the scatter plots that the data is dispersed and the data shows no other clear tendency or pattern. Further investigations have found that none of the higher-order terms (squares or cubes of the predictors) improved the model. A linear relationship may be a target of criticism, but we found no other simple relationship more suitable for model building.

At the beginning of our investigations, predictor values were based on a series of CT slices. Even these predictor values may be affected by measurement errors. In this investigation, we did not have the opportunity to perform repeated measures or to set the predictor values to investigate the effect of predictor measurement errors. The use of the OLS (ordinary least squares) estimation (to estimate model coefficients) might be criticised, but the assumptions of the linear regression model were checked carefully and the residual plot demonstrated no connection between the variance of the dependent variable and the predictor values. Residual plots are also of great importance to evaluate the goodness of fit of the model used [39].

Predictor values based on only one CT slice are measured with supposedly larger errors. At this stage, the predictors were measured both ways (i.e. based on a series of CT slices and based on only one CT slice). This design allowed us to evaluate the predictor errors based on only one CT slice and to evaluate the consequences in a sample of 100 patients.

There are certain limitations of the linear regression model. The performance of the model is fair, but limited to a sensitivity of 80.7% and a specificity of 87.5%. These values seemed to be stable throughout the different steps of the evaluation [26].

Comparison with other studies in the context of other heart-sparing methods

There is a strong effort to minimize OAR doses during radiation therapy.

As an initial approach, in Lymberis et al., a Spearman correlation analysis was applied to identify a monotonic relationship between OAR dose values and in-field OAR volumes [40]. At present, the risk of radiogenic heart sequelae is thought to be related to the mean dose to the heart [41, 42]. Based on epidemiological data and the simulation of out-of-date techniques, the mean heart dose is a good approximation; however, many centres also focus on the dose to the coronary arteries, most significantly the LAD. We believe that the LAD dose is a rational measure of the danger to the heart in left-sided cases [41, 43-45].

Our LAD dose constraints (i.e. 12.5 Gy and 12.9 Gy in the supine and prone positions, respectively) can be regarded as low and strict dose values compared to Jacob et al. [46].

Zhao at el. built an SVM (support vector machine)-based two-step decision-making algorithm in a sample of 198 patients [47]. Their method is based on anatomical characteristics measured on a prone CT series. The outcome of that test decides whether the patient benefits from the prone position radiotherapy or should undergo a CT series in the supine position for comparison. Although numerical measures of the goodness of classification were impressive, that tool provided no numerical estimate of the advantage of one treatment position over the other; hence, no optimization could be practised. With the SVM method, the selection of cases with an in-field heart volume over the acceptable threshold in the prone position could be managed, and in some cases, a second CT series was needed in the supine position. Furthermore, no volumetric dose data could be collected [47].

Another potential method to minimize OAR dose values is the use of the deep inspiration breath hold technique (DIBH). Lin et al. compared the DIBH method and prone positioning [48]. Their results demonstrate that it is rather difficult to distinguish patients who benefit from DIBH versus patients who have benefited from prone positioning. Although the mathematical and IT apparatus they used is impressive, their results were based on a relatively small sample of 16 patients and therefore their results might be rather limited.

Recently, the optimal use of prone and supine positioning vs. other kinds of techniques (such as DIBH and IMRT) has become a topic of intensive research. The results of Lin et. al suggest that the classification problem (that is, predicting the best intervention method) can be statistically very challenging [48].

Recommendations for clinical practice

Although this thesis has discussed the use of the LAD mean dose difference in detail as a primary outcome, there are other possible dependent variables, such as the heart dose. In clinical practice, an algorithm for the dose to the heart is also considered in a complex decision, but only as a secondary outcome measure [26]. In most centres, the most widely applied outcome measure is the mean heart dose [42, 49]. Since a linear, non-threshold association exists between the mean heart dose and coronary events, the mean heart dose may be regarded as an approximation of the doses to the LAD and other coronary arteries [49].

Since there is a strong correlation between mean heart dose and LAD dose (R=0.87 in both positions), we believe that the predictive model presented here could be adapted to local practice after careful investigation in any centre applying prone radiotherapy.

Moreover, the simple tool, which uses a single CT slice in the supine position, combined with the dose constraints described in [26] is in everyday use at the Department of Oncotherapy, University of Szeged. The linear regression-based model was also tested in a 28-case external dataset of left-side breast cancer patients from Liège and showed great consistency in our results noted above. Predicted treatment position was correct in 24 out of 28 (accuracy: 85.7%) cases [26].

5. Summary and new results

Decision curves proved to be useful in comparing our models. We consider the linear regression model as a useful tool to support physicists and physicians in predicting the preferable treatment position. These results are promising and show that these predictive models can be used effectively in everyday clinical practice with the dose constraints described. The linear regression-based calculator is in use at the University of Szeged Department of Oncotherapy.

- Our results identified a relation between the predictors (BMI, median distance and area) and the LAD mean dose difference.
- The multiple linear regression-based model was developed and has proved to be an effective prediction model for the selection of the optimal radiation therapy position of left breast cancer patients.
- It was shown that the application of decision curve analysis for ‘non-probability outcomes’ is effective. Our simulation results confirm the use of logistic regression to produce probabilities to calculate ‘net benefit’.
- The novel decision curve method has been effectively applied to evaluate the performance of this multiple linear regression model.
- The decision curve analysis and the random cross validation method clearly present the effectiveness of the multiple linear regression model.
- The multiple regression-based model is applied in everyday medical practice (under well-defined dose constraints) at the University of Szeged Department of Oncotherapy [26]. The multiple regression-based model is applicable to predict preferable treatment position.

Acknowledgements

I am greatly obliged to Dr. Krisztina Boda and to Professor Tibor Nyári, my supervisors, for their continuous support and guidance in my scientific activities.

I am deeply indebted to Professor Zsuzsanna Kahán, former director of the Department of Oncotherapy, for her continuous optimism and immensely supportive encouragement.

I am very grateful to Dr. Renáta Lilla Kószó and Dr. Zoltán Varga (Department of Oncotherapy) for providing me with the opportunity to collaborate with them.

I also wish to thank all my co-authors for the chance to work with them.

I would like to express my gratitude to Professor Ferenc Bari and Professor Ferenc Peták, former and current heads of the Department of Medical Physics and Informatics, for providing me with the facilities necessary to complete this thesis and for their continuous kind support.

Thanks as well to all my colleagues at the Department of Medical Physics for being so friendly and supportive.

I am very grateful to Dr. Thomas Williams for his thoroughgoing work in reviewing the language of the thesis.

I would like to express my appreciation to my teachers and professors in the past for their conscientious work in educating me.

I wish to thank all my friends for their kind, supportive attitude.

I am especially thankful to my parents and my family for their never-ending supportive attitude and for their enormous encouragement.

With this dissertation, I would like to thank my parents in particular for their emotional and financial support throughout my studies and for the fact that they have always believed in me.

This research was partly supported by the European Union and the State of Hungary, co-financed by the European Social Fund within the framework of EFOP-3.6.1-16-2016-00008.

References

1. Tripepi, G., et al., *How to deal with continuous and dichotomic outcomes in epidemiological research: linear and logistic regression analyses.* Nephron Clin Pract, 2011. **118**(4): p. c399-406.
2. Vickers, A.J. and E.B. Elkin, *Decision curve analysis: a novel method for evaluating prediction models.* Med Decis Making, 2006. **26**(6): p. 565-74.
3. Kemény, S., et al., *Kísérletek tervezése és értékelése*. 2017, Budapest: Typotex.
4. Rawlings, J.O., S.G. Pantula, and D.A. Dickey, *Applied regression analysis : a research tool*. 2nd ed. Springer texts in statistics. 1998, New York: Springer. xviii, 657 p.
5. Draper, N.R. and H. Smith, *Applied regression analysis*. 3rd ed. Wiley series in probability and statistics Texts and references section. 1998, New York: Wiley. xvii, 706 p.
6. Montgomery, D.C. and G.C. Runger, *Applied statistics and probability for engineers*. 3rd ed. 2003, New York: Wiley. xiv, 706 p.
7. Stoltzfus, J.C., *Logistic regression: a brief primer.* Acad Emerg Med, 2011. **18**(10): p. 1099-104.
8. Hosmer, D.W. and S. Lemeshow, *Applied logistic regression*. 2nd ed. Wiley series in probability and statistics Texts and references section. 2000, New York: Wiley. xii, 373 p.
9. Altman, D.G. and P. Royston, *What do we mean by validating a prognostic model?* Stat Med, 2000. **19**(4): p. 453-73.
10. Agresti, A., *Categorical data analysis*. 2nd ed. Wiley series in probability and statistics. 2002, New York: Wiley-Interscience. xv, 710 p.
11. Baker, S.G. and B.S. Kramer, *Peirce, Youden, and Receiver Operating Characteristic Curves.* The American Statistician, 2007. **61**(4): p. 343-346.
12. López-Ratón, M., et al., *OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests.* Journal of Statistical Software; Vol 1, Issue 8 (2014), 2014.
13. Obuchowski, N.A., M.L. Lieber, and F.H. Wians, Jr., *ROC curves in clinical chemistry: uses, misuses, and possible solutions.* Clin Chem, 2004. **50**(7): p. 1118-25.
14. Steyerberg, E.W., et al., *Graphical assessment of incremental value of novel markers in prediction models: From statistical to decision analytical perspectives.* Biom J, 2015. **57**(4): p. 556-70.
15. BRIER, G.W., *VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY.* Monthly Weather Review, 1950. **78**(1): p. 1-3.
16. Augustin, H., et al., *Decision curve analysis to compare 3 versions of Partin Tables to predict final pathologic stage.* Urol Oncol, 2012. **30**(4): p. 396-401.
17. Collins, G.S. and D.G. Altman, *Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2.* BMJ, 2012. **344**: p. e4181.

18. Godoy, G., et al., *Extent of pelvic lymph node dissection and the impact of standard template dissection on nomogram prediction of lymph node involvement.* Eur Urol, 2011. **60**(2): p. 195-201.

19. Lughezzani, G., et al., *Comparison of three different tools for prediction of seminal vesicle invasion at radical prostatectomy.* Eur Urol, 2012. **62**(4): p. 590-6.

20. Steyerberg, E.W., et al., *Assessing the performance of prediction models: a framework for traditional and novel measures.* Epidemiology, 2010. **21**(1): p. 128-38.

21. Ankerst, D.P., et al., *Updating risk prediction tools: a case study in prostate cancer.* Biom J, 2012. **54**(1): p. 127-42.

22. Barbieri, C.E., et al., *Decision curve analysis assessing the clinical benefit of NMP22 in the detection of bladder cancer: secondary analysis of a prospective trial.* BJU Int, 2012. **109**(5): p. 685-90.

23. Peirce, C.S., *The numerical measure of the success of predictions.* Science, 1884. **4**(93): p. 453-4.

24. Varga, Z., et al., *Individualized positioning for maximum heart protection during breast irradiation.* Acta Oncol, 2014. **53**(1): p. 58-64.

25. Rarosi, F., et al., *Decision curve analysis apropos of choice of preferable treatment positioning during breast irradiation.* BMC Med Inform Decis Mak, 2019. **19**(1): p. 204.

26. Kahan, Z., et al., *A simple clinical method for predicting the benefit of prone vs. supine positioning in reducing heart exposure during left breast radiotherapy.* Radiother Oncol, 2018. **126**(3): p. 487-492.

27. Bland, J.M. and D.G. Altman, *Statistical methods for assessing agreement between two methods of clinical measurement.* Lancet, 1986. **1**(8476): p. 307-10.

28. Kolmogorov, A.N., *Sulla determinazione empirica di una legge di distribuzione.* Giorn. Ist. Ital. Attuari, 1933(4): p. 83-91.

29. Kolmogorov, A.N., *Confidence Limits for an Unknown Distribution Function.* Ann. Math. Statist., 1941. **12**(4): p. 461-463.

30. Smirnov, N.V., *Tables for estimating the goodness of fit of empirical distributions.* Annals of Mathematical Statistics, 1948. **19**: p. 279-281.

31. Durbin, J. and G.S. Watson, *Testing for serial correlation in least squares regression. I.* Biometrika, 1950. **37**(3-4): p. 409-28.

32. Durbin, J. and G.S. Watson, *Testing for serial correlation in least squares regression. II.* Biometrika, 1951. **38**(1-2): p. 159-78.

33. Vickers, A.J. *Decision curve analysis*. [homepage] 2010; [program packages]. Available from: www.decisioncurveanalysis.org.

34. Kószó, R.L., *Individualised radiotherapy serving reducedtoxicityin breast and prostate cancer*, in *Doctoral School of Interdisciplinary Sciences, Faculty of Medicine, University of Szeged, Hungary*. 2018, University of Szeged: Szeged.

35. Bland, J.M. and D.J. Altman, *Regression analysis.* Lancet, 1986. **1**(8486): p. 908-9.

36. Kasza, J. and R. Wolfe, *Interpretation of commonly used statistical regression models.* Respirology, 2014. **19**(1): p. 14-21.

37. Lee, K.J., M.M. Wiest, and J.B. Carlin, *Statistics for clinicians: an introduction to linear regression.* J Paediatr Child Health, 2014. **50**(12): p. 940-3.

38. Schneider, A., G. Hommel, and M. Blettner, *Linear regression analysis: part 14 of a series on evaluation of scientific publications.* Dtsch Arztebl Int, 2010. **107**(44): p. 776-82.

39. Godfrey, K., *Simple linear regression in medical research.* N Engl J Med, 1985. **313**(26): p. 1629-36.

40. Lymberis, S.C., et al., *Prospective assessment of optimal individual position (prone versus supine) for breast radiotherapy: volumetric and dosimetric correlations in 100 patients.* Int J Radiat Oncol Biol Phys, 2012. **84**(4): p. 902-9.

41. Darby, S.C., et al., *Risk of ischemic heart disease in women after radiotherapy for breast cancer.* N Engl J Med, 2013. **368**(11): p. 987-98.

42. Kirby, A.M., et al., *Prone versus supine positioning for whole and partial-breast radiotherapy: a comparison of non-target tissue dosimetry.* Radiother Oncol, 2010. **96**(2): p. 178-84.

43. Taylor, C.W. and A.M. Kirby, *Cardiac Side-effects From Breast Cancer Radiotherapy.* Clin Oncol (R Coll Radiol), 2015. **27**(11): p. 621-9.

44. Taylor, C.W., et al., *Exposure of the Heart in Breast Cancer Radiation Therapy: A Systematic Review of Heart Doses Published During 2003 to 2013.* Int J Radiat Oncol Biol Phys, 2015. **93**(4): p. 845-53.

45. Duane, F.K., et al., *Cardiac Structure Doses in Women Irradiated for Breast Cancer in the Past and Their Use in Epidemiological Studies.* Pract Radiat Oncol, 2019. **9**(3): p. 158-171.

46. Jacob, S., et al., *Is mean heart dose a relevant surrogate parameter of left ventricle and coronary arteries exposure during breast cancer radiotherapy: a dosimetric evaluation based on individually-determined radiation dose (BACCARAT study).* Radiat Oncol, 2019. **14**(1): p. 29.

47. Zhao, X., et al., *A support vector machine (SVM) for predicting preferred treatment position in radiotherapy of patients with breast cancer.* Med Phys, 2010. **37**(10): p. 5341-50.

48. Lin, H., et al., *Feasibility study of individualized optimal positioning selection for left-sided whole breast radiotherapy: DIBH or prone.* J Appl Clin Med Phys, 2018. **19**(2): p. 218-229.

49. Wurschmidt, F., et al., *Incidental dose to coronary arteries is higher in prone than in supine whole breast irradiation. A dosimetric comparison in adjuvant radiotherapy of early stage breast cancer.* Strahlenther Onkol, 2014. **190**(6): p. 563-8.

1.sz. melléklet

Társszerzői lemondó nyilatkozat

Alulírott Prof. Dr. Kahán Zsuzsanna (felelős társszerző) kijelentem, hogy Rárosi Ferenc (pályázó) Ph.D. értekezésének tézispontjaiban bemutatott - közösen publikált - tudományos eredmények elérésében a pályázónak meghatározó szerepe volt, ezért ezeket a téziseket más a PhD fokozat megszerzését célzó minősítési eljárásban nem használta fel, illetve nem kívánja felhasználni.

Szeged
2020.VI.8.

.....
Prof. Dr. Kahán Zsuzsanna

A pályázó tézispontjaiban érintett, közösen publikált közlemények:

Rárosi, F., Boda, K., Kahán, Z., and Varga, Z. (2019). Decision curve analysis apropos of choice of preferable treatment positioning during breast irradiation. BMC Medical Informatics and Decision Making 19(1):204.

Varga, Z., Cserháti, A., Rárosi, F., Boda, K., Gulyás, G., Együd, Z., and Kahán, Z. (2014). Individualized positioning for maximum heart protection during breast irradiation. Acta Oncologica 53, 58–64.

TÉMAVEZETŐI NYILATKOZAT

Kijelentem, hogy Dr. Kószó Renáta Lilla doktorjelölt „Individualised radiotherapy serving reduced toxicity in breast and prostate cancer" című doktori értekezésében jelentősek Rárosi Ferenc doktorjelölt érdemei. A statisztikai analízis és az adatok interpretációja tükrözi a munkához való hozzájárulását.

Vonatkozó közös benyújtott kézirat: Kószó R., Kahán Z., Darázs B., Rárosi F., Varga Z. Dosimetric comparison of 3D-CRT, sliding window IMRT and VMAT techniques for external beam accelerated partial breast irradiation. Acta Oncologica.

Vonatkozó közös megjelent publikáció: Kahán Z., Rárosi F., Gaál S., Cserháti A., Boda K., Darázs B., **Kószó R.**, Lakosi F., Gulyban Á., Coucke PA., Varga Z. A simple clinical method for predicting the benefit of prone vs. supine positioning in reducing heart exposure during left breast radiotherapy. Radiother Oncol. 2018; 126: 487–492.

Kelt: Szeged, 2018.12.17.

.....
Dr. Varga Zoltán

.....
Stéhlikné Dr. Boda Krisztina

.....
Dr. Nyári Tibor
(Rárosi Ferenc témavezetői)

(Dr. Kószó Renáta Lilla témavezetője)

I. **Rárosi, F.**, Boda, K., Kahán, Z., and Varga, Z. (2019). Decision curve analysis apropos of choice of preferable treatment positioning during breast irradiation. BMC Medical Informatics and Decision Making 19(1), 204. IF: 2.067

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

Decision curve analysis apropos of choice of preferable treatment positioning during breast irradiation

Ferenc Rárosi[1*], Krisztina Boda[1], Zsuzsanna Kahán[2] and Zoltán Varga[2]

Abstract

Background: Radiotherapy is a standard treatment option for breast cancer, but it may lead to significant late morbidity, including radiation heart damage. Breast irradiation performed individually in the supine or prone position may aid in minimizing the irradiation dose to the heart and LAD coronary artery. A series of CT scans and therapy plans are needed in both positions for the 'gold standard' decision on the preferable treatment position. This method is expensive with respect to technology and physician workload.
Our ultimate goal is to develop a predictive tool to identify the preferable treatment position using easily measurable patient characteristics. In this article, we describe the details of how model building and consequently validation of the best model are done.

Methods: Different models were used: both logistic regression and multiple linear regressions were used to estimate the LAD mean dose difference (the difference between the mean dose to the LAD in the supine position versus prone position); predicted dose differences were analysed compared to the 'gold standard' values, and the best model was selected accordingly. The final model was checked by random cross-validation. In addition to generally used measures (ROC and Brier score), decision curves were employed to evaluate the performance of the models.

Results: ROC analysis demonstrated that none of the predictors alone was satisfactory. Multiple logistic regression models and the linear regression model lead to high values of net benefit for a wide range of threshold probabilities. Multiple linear regression seemed to be the most useful model. We also present the results of the random cross-validation for this model (i.e. sensitivity of 80.7% and specificity of 87.5%).

Conclusions: Decision curves proved to be useful to evaluate our models. Our results indicate that any of the models could be implemented in clinical practice, but the linear regression model is the most useful model to facilitate the radiation treatment decision. In addition, it is in use in everyday practice in the Department of Oncotherapy, University of Szeged, Hungary.

Keywords: Decision curve, Regression model, Prediction, Validation, Left-sided breast radiotherapy, LAD mean dose

* Correspondence: rarosi.ferenc@med.u-szeged.hu
[1]Department of Medical Physics and Informatics, University of Szeged, Korányi fasor 9, Szeged 6720, Hungary
Full list of author information is available at the end of the article

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

Background

Radiotherapy is an effective treatment for breast cancer, but it can lead to significant late morbidity, particularly connected to various heart diseases [1, 2]. The goal of radiotherapy is to achieve a good therapeutic ratio, i.e. a sufficient dose in the planning target volume but as low as possible to normal tissues and especially to critical organs at risk (OAR). Great efforts are made to minimize the irradiation dose to the heart in general; nevertheless, the dose to the LAD is of the utmost clinical importance [3]. There are different approaches to protect the heart from radiation, among which breathing control, including the deep inspirational breast hold technique and IMRT or partial breast irradiation, have just recently become available in routine practice [4, 5]. Originally, prone positioning was invented for the irradiation of difficult cases with large breasts, and many groups came to favour its wide use [6–11]. Numerous studies including ours have demonstrated that the preferable position varies from patient to patient [6–8, 12–16]. Some investigations have identified patient-related predictors in favour of a particular treatment position, while others have not.

There are studies that prefer the prone position for heart protection in general; others did not find a relevant benefit of prone positioning compared to supine positioning [8, 16]. Investigations have demonstrated that the favourable treatment set-up varies from patient to patient but have not shown patient-related characteristics to predict the favourable treatment position [8, 9, 15, 16]. Others have shown an association between breast size and the benefit of prone positioning [8]. In our prospective clinical experiment, we found a strong relationship between the dose to the LAD and heart and some patient-related characteristics, such as BMI and the geography of the breast, heart and chest wall [6, 7].

Prediction models are widely used in biomedical research and other interdisciplinary fields of research. These models are mainly based on a regression method: if the dependent variable is continuous, then a multiple regression model can be used, while logistic regression is applied for categorical (often binary) dependent variables. The result of a regression model is an expected value of the dependent variable, and the result of a (binary) logistic regression is an expected probability. When the purpose is to make a decision and make predictions concerning the existence of a phenomenon, such as an illness or the necessity of an operation, the decision is based on a carefully chosen cut-point.

There are several measures to describe the performance of the prediction model, with the ROC curve being the most commonly used one [17, 18]. Unfortunately, it has the disadvantage of equally weighting false positive and false negative decisions, whereas it may be important to weight the different types of misclassifications differently.

Vickers has published a method which not only eliminates this weakness in the ROC method, but also introduces a completely different approach to derive a measure called ‘net benefit’ to evaluate the clinical utility of the method [19]. The value of the net benefit depends on the cut-point chosen. We can obtain the decision curve if we plot the net benefit in function of the cut-point (threshold probability). Decision curves can be effectively used to compare different prediction methods and to determine the range of the possible cut-point, where the use of the prediction model is beneficial. We report on how the decision curve method has been applied to evaluate different models to facilitate individualized breast irradiation.

A series of CT scans and therapy plans in both positions (supine and prone) are called for to compare the dose to the heart and LAD and to select the preferable treatment position; we call this optimization process the ‘gold standard’ decision. This method is expensive (with respect to both the technology and physician workload) and involves an extra dose of radiation to the patients. This was the motivation for creating a model to predict the preferable position and anticipate the dose difference between the two positions using the patient-related characteristics noted above. In fact, this model is already being implemented in routine radiotherapy practice at the Department of Oncotherapy, University of Szeged, Szeged, Hungary; in this article, we describe the details of how model building and consequently validation of the best model took place [6, 7].

Methods

Description of data

Various patient-related features and dosimetry data extracted from radiation treatment plans generated in both prone and supine positions of 83 left-sided breast cancer cases receiving postoperative whole-breast radiotherapy were used for analyses. The details of the IRB-approved clinical study have been described elsewhere [6]. LAD and heart doses were related most strongly to body mass index (BMI) and the geography of the heart and breast. The latter could be characterized by the median distance between the LAD and the chest wall (d_{median}), and the heart area included in the radiation field on a single CT scan at the middle of the heart in the supine position (A_{heart}). The effect of some other measures, such as the traditionally considered breast size (corresponding to PTV), waist and hip circumferences, were less important [6]. In another 55 cases, these preliminary observations showed great consistency. Detailed evaluation was carried out on a total of 138 cases in this study. Although both the LAD and heart doses were considered in the model, since the dose to the LAD has the greatest impact on clinical decision making, this parameter is presented as the only dependent variable here.

We considered the prone position preferable if the LAD dose was smaller in that position than in the supine position, and we defined the supine position as preferable if the opposite was true. The dependent variable was the difference in the mean dose to the LAD ('LAD mean dose difference'), meaning the mean dose to the LAD in the supine position minus the mean dose to the LAD in the prone position as derived from the radiation treatment planning system. In the event of a positive LAD mean dose difference, the prone position seemed more favourable, while the supine position was more advantageous in the event of a negative LAD mean dose difference. The LAD mean dose difference is a continuous dependent variable, but the decision is binary. The 'gold standard decision' on the treatment position is based on the value of the LAD mean dose difference.

Prediction models

Our early models were logistic regression-based models using patients' characteristics noted in the 'Description of data' section. The continuous dependent variable had to be dichotomized for the logistic regression-based models. We coded it 0 if the prone position was preferred and 1 if the supine position was preferred, based on the gold standard method discussed in detail above. A hierarchical cluster analysis was performed with a similarity measure of the Pearson correlation coefficient to avoid multicollinearity between independent variables. The correlations were explored between the predictors, with poorly correlated predictors chosen for model building. We were searching for a parsimonious model with relatively few and uncorrelated predictors. We applied the backward and forward likelihood ratio selection methods to select a logistic regression model.

The dependent variable was originally continuous in our dataset; the use of binary logistic regression would have led to a loss of information. Treating the dependent variable as a continuous variable, we used a multiple linear regression. Higher differences are taken into account with more weight in the regression model. Another advantage of using multiple linear regression is that the expected value of the LAD mean dose difference can also be calculated. The decision was based on the sign of the estimated dependent variable in the regression model.

It is well known that the results of the classifications are overly optimistic when the classification is made on the same dataset where the classification rule was discovered (i.e. the training set and the test set are the same). To identify the best model, we used an internal validation method of data splitting as follows: the sample was divided into two parts randomly, with 70% of the sample as the training set for the linear regression and the resultant model being tested on the remaining 30% of the data. The classification results and the misclassified dose were also noted. The process was repeated 1000 times randomly, and it was not just the proportions of misclassified patients that were taken into account, but also the distribution of the misclassified dose.

The calculations were carried out in IBM SPSS version 24 and R (version 3.3.1).

ROC analysis

The performance of a prediction model can be evaluated by comparing the decision to the gold standard method. In most cases, we simply do not know the 'truth'. The nearest we have to it is the gold standard, so we have to regard that as the 'truth'. There are several measures to describe the performance of the prediction model based on the numbers in TP (true positive), FP (false positive), TN (true negative) and FN (false negative) cases. The well-known measures are sensitivity, specificity, positive predictive value, negative predictive value, accuracy (proportion of all correct diagnoses) and the Youden index (i.e. sensitivity+specificity-1). The method involving ROC curves is based on the measures noted above: we plot the sensitivity in function of 1-specificity at various threshold levels. We can also choose the optimal cut-point based on these measures and methods. One of the most frequently used methods for cut-point selection is to find the maximum value of the Youden index, with the value that maximizes the Youden index being the cut-point. Other methods for choosing a cut-point have been published [17, 18, 20–22].

Brier score

The Brier score, originally introduced by Glenn W. Brier, is calculated as follows:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2 \tag{1}$$

and the formula for the weighted Brier score is:

$$wBS = \frac{\sum_{i=1}^{N} w_i(p_i - o_i)^2}{\sum_{i=1}^{N} w_i} \tag{2}$$

where o_i is the i^{th} outcome, p_i is the i^{th} predicted probability and w_i is the weight of the i^{th} item. If the outcome occurs in the i^{th} case, then $o_i = 1$; otherwise, $o_i = 0$. The Brier score measures an average (or weighted average) of squared distances between current outcomes and predicted probabilities, while lower values indicate better predictive performance [23].

The performance of the three best predictors (BMI, area and median distance) was evaluated by univariate logistic regression. For example, univariate logistic regression was

employed with BMI as the independent variable (and the preferable treatment position as the dependent variable) to construct a probability prediction based on the value of BMI.

Net benefit and decision curves

Decision curve analysis is a relatively new method to evaluate the performance of diagnostic tests and prediction models. It is based on the predicted probabilities of statistical models. The decision curve method was introduced by Vickers, based on the 'net benefit function' [19]. A decision curve is a plot which shows the net benefit calculated at various threshold levels. The definition of net benefit is based on the 'utility of the prediction of the method', originally defined by Peirce as:

$$B = \frac{p \cdot TP - l \cdot FP}{TP + FP + FN + TN} = \frac{p \cdot TP - l \cdot FP}{N} \tag{3}$$

where p stands for the 'profit' of a true positive decision, l refers to the 'loss' of a false positive decision, TP, FP, TN and FN are the number of true positive, false positive, true negative and false negative decisions, respectively, and N is the sample size [24].

The net benefit is defined as the benefit divided by the profit:

$$NB = \frac{B}{p} = \frac{TP}{N} - \frac{l}{p}\frac{FP}{N} \tag{4}$$

In other words, the net benefit is the benefit that results from the normalization of the profit. In this aspect, the 'loss-to-profit ratio' is a weighting factor to give weight to the false positive decision compared to one unit of benefit of the true positive decision. It is important to note that profit and loss are unknown in most applications and it is impossible to measure them. This is common in medical decision making. One simply cannot measure or numerically anticipate the consequences of the true positive decision (the so-called profit of the operation, for instance) or the consequences of the false positive decision (for example, the loss of an unnecessary operation). That is why Vickers and Elkin suggested calculating this weighting factor (loss-to-profit ratio) as the odds of the threshold probability. The weights of the four possible outcomes (TP, FP, FN and TN) are not known; still, it is possible to make an acceptable assumption of the 'loss'-to-'profit' ratio.

There is an important assumption by Vickers:

$$\frac{l}{p} := \frac{p_t}{1 - p_t} \tag{5}$$

where p_t is the threshold probability (or cut-point), above which the outcome of a probability prediction model is labelled 'positive' and below which it is labelled 'negative' [22].

If we accept this assumption, net benefit simplifies to:

$$NB = \frac{B}{p} = \frac{TP}{N} - \frac{p_t}{1 - p_t}\frac{FP}{N} \tag{6}$$

The decision curve is a curve that illustrates the net benefit in function of the threshold level $\boldsymbol{p_t}$. Using this method, we can compare the performance of different predictive models to show which model is more beneficial in function of the threshold probability. This method also shows the range of threshold levels where the decision is beneficial.

Results

Primary results for the predictors

Our investigations revealed that none of the predictors alone was sufficient for prediction. Candidate predictor PTV had AUC-ROC of 0.722 while AUC-ROC for BMI was 0.740, which is fair, but not sufficient for our purposes.

The best predictor was A_{heart}, with AUC-ROC of 0.868. Table 1 shows the classification results of the most important candidate predictors and the primary results of the multivariate prediction models. These values in Table 1 are not cross-validated and based on data from 83 patients. Table 1 presents the model selection and the performance of the predictors alone.

Although it is possible to base recommendations on a single predictor value, our experience suggests that the

Table 1 Classification results for the predictors and for the multivariate prediction models based on data from $n = 83$

Model	ROC-AUC	95% Confidence interval for ROC-AUC	Brier score
PTV (PTV candidate predictor only)	0.722	0.608, 0.836	0.213
BMI (BMI predictor only)	0.740	0.630, 0.850	0.201
Median distance (d_{median} predictor only)	0.787	0.690, 0.884	0.189
Area (A_{heart} predictor only)	0.868	0.791, 0.944	0.151
Logistic regression (main effect model, Model1)	0.906	0.854, 0.959	0.124
Logistic regression (forward LR selection model, Model2)	0.900	0.848, 0.953	0.132
Linear regression (Model3)	0.903	0.850, 0.957	0.139

predictive power of this simple approach is not satisfactory. This is why we constructed a multivariate prediction model.

Multivariate models

There are several complex predictive methods in statistics, but our goal was to use a simple model with acceptable predictive performance.

Among the multivariate models, the backward likelihood ratio selection model was the 'main effect model': *area + BMI + median distance*. The forward likelihood ratio selection model was not a hierarchical model, as it contained two interaction terms: *area*BMI + area*median distance* without main effects.

Multiple linear regression seemed to be the most useful model with respect to the promising high AUC-ROC of 0.903 (results in Table 1) and the advantage noted above that the expected value of the LAD mean dose difference can also be calculated.

The results of the random cross-validation for the multiple regression model have a sensitivity of 80.7% and specificity of 87.5% [6].

Results of comparing the models using decision curves

We examined the performance of the three best predictors alone with decision curves, and we compared them to the models described above. None of the best three predictors alone was comparable to the prediction models (Fig. 1).

The decision curves for the logistic regression model and the linear regression-based model were quite similar to each other. We can conclude that both models lead to high values of net benefit for a wide range of threshold probabilities. In other words, it is beneficial to use these models in respect of net benefit regardless of the current threshold probability. These results showed that these models can be used in clinical practice.

A bootstrap method was applied to construct 95% confidence intervals for the net benefit at various threshold levels for the main effect model [7]. 1000 bootstrap samples were generated by Microsoft Excel with a sample size of 138 each. Net benefit values were calculated by definition. The 2.5 and 97.5 percentiles of the net benefit values were calculated at each threshold level. The results are shown in Table 2.

The linear regression model is mathematically simple and can be very easily implemented (for instance, in Microsoft Excel). To the best of our knowledge, no similar composite models have been used to select treatment position in radiotherapy.

Discussion

We presented different models based on patient-related parameters to predict the preferable treatment position in left-sided breast cancer radiotherapy and the mathematical aspects of the evaluation of the predictive power of these models. All the predictive models performed better than single predictors did, but the linear regression model

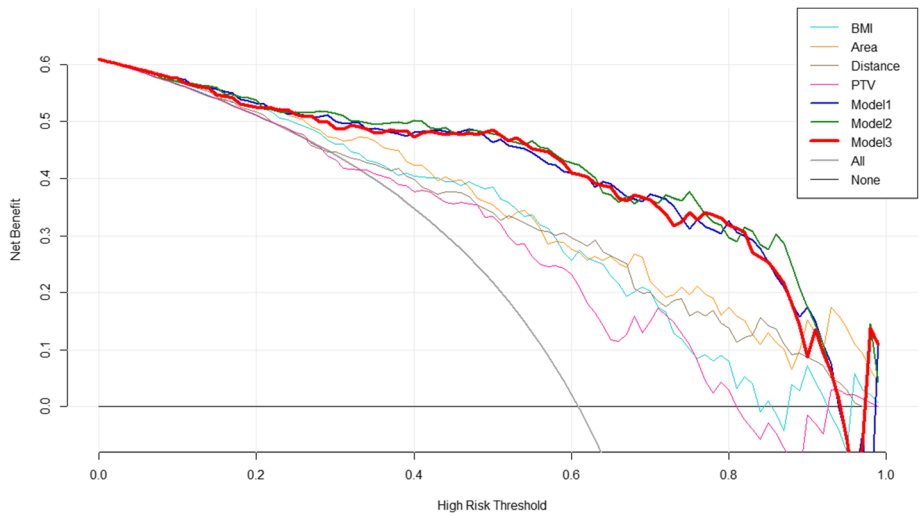


Fig. 1 Decision curves for the four best predictors and three multivariate models. The vertical axis represents the value of net benefit, and the horizontal axis represents the threshold level (possible probability cut-points). Plotting net benefit in function of threshold level yields the decision curve. In the legend, PTV, Area, Distance and BMI refer to candidate predictor PTV, predictor A_{heart}, predictor d_{median} and predictor BMI alone, respectively. Model1 to Model3 refer to the performance of the multivariate prediction models. Model2 is the main effect model (*area + BMI + median distance*), Model2 is the forward likelihood ratio selection model with two interaction terms (*area*BMI + area*median distance*), and Model3 is the linear regression-based model. This figure shows that the logistic regression models and the linear regression model lead to very similar high values of net benefit in a wide range of threshold levels and that none of the predictors alone can lead to similarly high values of net benefit

Table 2 95% confidence intervals for net benefit (logistic regression, main effect model) at different threshold levels. Confidence bounds are based on 1000 bootstrap samples

Threshold probability	Lower bound for net benefit	Upper bound for net benefit
0.1	0.564	0.573
0.2	0.496	0.547
0.3	0.450	0.527
0.4	0.423	0.517
0.5	0.391	0.507
0.6	0.319	0.467
0.7	0.256	0.435
0.8	0.210	0.435
0.9	0.000	0.333

was considered the most clinically relevant for quantitative estimation.

One merit of our study is its relatively large sample size and multivariate aspect. Decision curves and the AUC for the ROC results were found to be similar for the linear regression model and the two logistic regression models. Nevertheless, while the logistic regression models weight the outcomes on a binary scale, the linear regression model weights them in keeping with the magnitude of the difference. Since the linear regression model provides additional information, i.e. estimation of the dose difference, we decided to use the linear regression model. Knowledge of the estimated quantitative benefit of one or the other treatment position during radiotherapy may provide better guidance for the physician when considering various aspects, such as repositioning accuracy, patient comfort etc.

The application of AUC-ROC and measures like sensitivity and specificity is very common in radiotherapy planning, but we have not seen the approach of using decision curves in this field. Our investigations point to the clinical utility of predictive models.

There are certain limitations of the linear regression model we presented. The performance of the model is fair, but limited to a sensitivity of 80.7% and a specificity of 87.5%. These values seemed very stable throughout the different steps of the evaluation. Furthermore, in the next phase of development, very similar results were also found. In brief, a simple clinical tool which used the model was created and tested for clinical practice [7]. This tool estimates the difference of the expected dose values based on the BMI and the d_{median} and A_{heart} measured on a CT slice at the middle of the heart. The result was compared to that of the full CT series in both positions and the dosimetric data. The comparison revealed very consistent results from the simple tool and the original method (very similar sensitivity and specificity values) [7].

One limitation might be that we assumed a linear relationship between the predictors and the dependent variable. Our investigation revealed that none of the higher-order terms (squares or cubes of the predictors) improved the model at all. In other words, the linear relationship may be a target of criticism, but we found no other simple relationship more suitable for model building.

Zhao at el. built an SVM (support vector machine)-based two-step decision algorithm in a sample of 198 patients [11]. Their method is based on anatomical characteristics measured on a prone CT series. This classified patients into prone position radiotherapy or into another CT series in supine position for comparison. Although the numerical measures of the goodness of classification were impressive, that tool provided no numerical estimation of the advantage of one treatment position over the other; hence, no optimization could be practised. With their method, one only manage to filter out cases with an in-field heart volume over the acceptable threshold in the prone position and with the necessity of a second CT series in the supine position [11].

As noted earlier, although this report only discussed the use of the LAD mean dose difference as a primary outcome, there are other possible dependent variables, such as the heart dose. In our clinical practice, algorithm of the dose to the heart is also considered in a complex decision, but only as a secondary outcome measure [7]. In most centres, the most widely applied outcome measure is the mean heart dose [8, 25, 26]. Since a linear, non-threshold association exists between the mean heart dose and coronary events, mean heart dose may be regarded as an approximation of the doses to the LAD and other coronary arteries [25, 26]. Since there is a strong correlation between mean heart dose and LAD dose ($R = 0.87$ in both positions), we believe that the predictive model presented here could be adapted to local practice in any centre applying prone radiotherapy.

Moreover, the simple tool, which uses a single CT scan in the supine position, combined with the dose constraints described in everyday practice at the Department of Oncotherapy, University of Szeged, is satisfactory and of great utility. The linear regression-based model was also tested in a 28-case external dataset of left-side breast cancer patients from Liege and showed great consistency in our results noted above. Predicted treatment position was correct in 24 out of 28 (accuracy: 85.7%) cases [7].

Conclusions

Our study has demonstrated that decision curves are useful in comparing our models. Any of the models could be implemented in clinical practice, but the linear regression model is the most useful and stable in facilitating the radiation treatment decision. In addition, this

linear regression model is already implemented in everyday radiotherapy practice at the Department of Oncotherapy, University of Szeged, Szeged, Hungary.

Abbreviations

A_{heart}: Area of the heart at the middle of the LAD; AUC: Area under curve; BMI: Body mass index; BS: Brier score; d_{median}: Median distance between the LAD and the chest wall; FN: Number of false negative cases; FP: Number of false positive cases; LAD: Left anterior descending coronary artery; LR: Likelihood ratio; NB: Net benefit; OAR: Organ at risk; PTV: Planning target volume (operated breast with safety margin); ROC: Receiver operating characteristic; TN: Number of true negative cases; TP: Number of true positive cases; T_p: Threshold probability; wBS: Weighted Brier score

Acknowledgements

The authors would like to thank Dr. Thomas Williams (Department of English, University of Szeged) for his help in reviewing the language in the manuscript.

Authors' contributions

FR analysed the patient data and investigated the predictive power of the candidate predictors. FR developed the prediction models, took part in model selection, validated the selected model and applied the decision curve approach. KB supervised the statistical aspects of the work, suggested the decision curve approach and revised the validation process. ZV planned the radiotherapy treatment, collected data and introduced the candidate predictors. ZK designed and supervised the medical aspects of the study and substantively revised the medical aspects. Each author has approved the submitted version (and any substantially modified version that involves that author's contribution to the study). Each author has agreed both to be personally accountable for his or her own contributions and to ensure that questions related to the accuracy or integrity of any part of the study, even those in which the author was not personally involved, are appropriately investigated and resolved and the resolution documented in the literature.

Funding

Funding for this study was provided in part by the VKSZ 12-1-2013-0012 project and the European Union and the State of Hungary, co-financed by the European Social Fund under EFOP-3.6.1-16-2016-00008. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of the University of Szeged (registration number 185/2012), and all enrolled patients have provided their written informed consent to participate in the study.

Consent for publication

The manuscript does not contain any individual person's data in any form.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Medical Physics and Informatics, University of Szeged, Korányi fasor 9, Szeged 6720, Hungary. [2]Department of Oncotherapy, University of Szeged, Korányi fasor 12, Szeged 6720, Hungary.

Received: 31 January 2019 Accepted: 14 October 2019
Published online: 29 October 2019

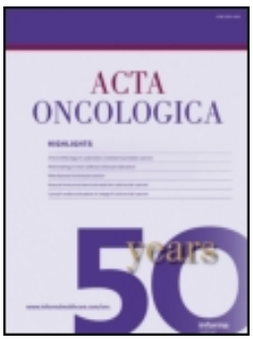
References

1. McGale P, Darby SC, Hall P, Adolfsson J, Bengtsson NO, Bennet AM, et al. Incidence of heart disease in 35,000 women treated with radiotherapy for breast cancer in Denmark and Sweden. Radiother Oncol. 2011;100:167–75.
2. Andratschke N, Maurer J, Molls M, Trott KR. Late radiation-induced heart disease after radiotherapy. Clinical importance, radiobiological mechanisms and strategies of prevention. Radiother Oncol. 2011;100:160–6.
3. Shah C, Badiyan S, Berry S, et al. Cardiac dose sparing and avoidance techniques in breast cancer radiotherapy. Radiother Oncol. 2014;112:9–16.
4. Taylor C.W.,-Kirby A.M. Cardiac Side-effects From Breast Cancer Radiotherapy. Clinicalm Oncology 2015; 27: 621–629.
5. Shah C, Tendulkar R, Smile T. el al. Adjuvant radiotherapy in early-stage breast Cancer: evidence-based options. Ann Surg Oncol. 2016;23:3880–90.
6. Varga Z, Cserhati A, Rarosi F, et al. Individualized positioning for maximum heart protection during breast irradiation. Acta Oncol. 2014;53(1):58–64.
7. Kahan Z, Rarosi F, Gaal S, et al. A simple clinical method for predicting the benefit of prone vs. supine positioning in reducing heart exposure during left breast radiotherapy. Radiother Oncol. 2018;126(3):487–92.
8. Kirby AM, Evans PM, Donovan EM, Convery HM, Haviland JS, Yarnold JR. Prone versus supine positioning for whole and partial-breast radiotherapy: a comparison of non-target tissue dosimetry. Radiother Oncol. 2010;96:178–84.
9. Formenti SC, Gidea-Addeo D, Goldberg JD, Roses DF, Guth A, Rosenstein BS, et al. Phase I–II trial of prone accelerated intensity modulated radiation therapy to the breast to optimally spare normal tissue. J Clin Oncol. 2007;25:2236–42.
10. Buijsen J, Jager JJ, Bovendeerd J, Voncken R, Borger JH, Boersma LJ, et al. Prone breast irradiation for pendulous breasts. Radiother Oncol. 2007;82:337–40.
11. Zhao X, Wong EK, Wang Y, Lymberis S, Wen B, Formenti S, et al. A support vector machine (SVM) for predicting preferred treatment position in radiotherapy of patients with breast cancer. Med Phys. 2010;37:5341–50.
12. Lymberis SC, Dewyngaert JK, Parhar P, Chhabra AM, Fenton-Kerimian M, Chang J, et al. Prospective assessment of optimal individual position (prone versus supine) for breast radiotherapy: volumetric and dosimetric correlations in 100 patients. Int J Radiat Oncol Biol Phys. 2012;84:902–9.
13. Formenti SC, DeWyngaert JK, Jozsef G, Goldberg JD. Prone vs supine positioning for breast cancer radiotherapy. JAMA. 2012;308:861–3.
14. Varga Z, Hideghéty K, Mező T, Nikolényi A, Thurzó L, Kahán Z. Individual positioning: a comparative study of adjuvant breast radiotherapy in the prone versus supine position. Int J Radiat Oncol Biol Phys. 2009;75:94–100.
15. Aznar MC, Korreman SS, Pedersen AN, Persson GF, Josipovic M, Specht L. Evaluation of dose to cardiac structures during breast irradiation. Br J Radiol. 2011;84:743–6.
16. Griem KL, Fetherston P, Kuznetsova M, Foster GS, Shott S, CHU J, et al. Three-dimensional photon dosimetry: a comparison of treatment of the intact breast in the supine and prone position. Int J Radiation Oncology Biol Phys. 2003;57:891–9.
17. Agresti A. Categorical data analysis. 2nd ed. New York: Wiley-Interscience; 2002.
18. Obuchowski NA, Lieber ML, Wians FH. ROC curves in clinical chemistry: uses, misuses, and possible solutions. Clin Chem. 2004;50(7):1118–25.
19. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak. 2006;26(6):565–74.
20. Baker SG, Kramer BS. Peirce, Youden, and receiver operating characteristic curves. Am Stat. 2007;61(4):343–6.
21. Lopez-Raton M, Cadarso-Suarez C, Rodriguez-Alvarez MX, Gude-Sampedro F. OptimalCutpoints: an R package for selecting optimal Cutpoints in diagnostic tests. J Stat Softw. 2014;61(8):1–36.
22. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models a framework for traditional and novel measures. Epidemiology. 2010;21(1):128–38.
23. Brier GW. Verification of forecasts expressed in terms of probability. Mon Wea Rev. 1950;78:1–3.
24. Peirce CS. The numerical measure of the success of predictions. Science. 1884;4:453–4.
25. Würschmidt F, Stoltenberg S, Kretschmer M, Petersen C. Incidental dose to coronary arteries is higher in prone than in supine whole breast irradiation. A dosimetric comparison in adjuvant radiotherapy of early stage breast cancer. Strahlenther Onkol. 2014;190:563–8.
26. Darby SC, Ewertz M, McGale P, et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. N Engl J Med. 2013;368:987–98.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

II. Varga, Z., Cserháti, A., **Rárosi, F.**, Boda, K., Gulyás, G., Együd, Z., and Kahán, Z. (2014). Individualized positioning for maximum heart protection during breast irradiation. Acta Oncologica 53, 58–64. IF: 2.997



Acta Oncologica

ISSN: 0284-186X (Print) 1651-226X (Online) Journal homepage: https://www.tandfonline.com/loi/ionc20

Individualized positioning for maximum heart protection during breast irradiation

Zoltán Varga, Adrienn Cserháti, Ferenc Rárosi, Krisztina Boda, Gergely Gulyás, Zsófia Együd & Zsuzsanna Kahán

To cite this article: Zoltán Varga, Adrienn Cserháti, Ferenc Rárosi, Krisztina Boda, Gergely Gulyás, Zsófia Együd & Zsuzsanna Kahán (2014) Individualized positioning for maximum heart protection during breast irradiation, Acta Oncologica, 53:1, 58-64, DOI: 10.3109/0284186X.2013.781674

To link to this article: https://doi.org/10.3109/0284186X.2013.781674

Published online: 02 Apr 2013.

Submit your article to this journal

Article views: 680

View related articles

View Crossmark data

Citing articles: 10 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ionc20

Acta Oncologica, 2014; 53: 58–64

informa
healthcare

ORIGINAL ARTICLE

Individualized positioning for maximum heart protection during breast irradiation

ZOLTÁN VARGA[1], ADRIENN CSERHÁTI[1], FERENC RÁROSI[2,3], KRISZTINA BODA[2], GERGELY GULYÁS[4], ZSÓFIA EGYÜD[1] & ZSUZSANNA KAHÁN[1]

[1]*Department of Oncotherapy, University of Szeged, Szeged, Hungary,* [2]*Department of Medical Physics and Informatics, University of Szeged, Szeged, Hungary,* [3]*Bolyai Institute, University of Szeged, Szeged, Hungary and* [4]*Institute of Informatics, University of Szeged, Szeged, Hungary*

Abstract

Background. Prone positioning has been found feasible and appropriate for the reduction of radiation exposure of the lungs, but its effects on the heart dose remain controversial. Individual anatomical features were sought for the selection of optimal treatment positioning. *Material and methods.* In 138 left-sided breast cancer cases awaiting postoperative whole-breast radiotherapy, conformal radiotherapy plans were generated in both prone and supine positions. *Results.* The radiation doses to the left anterior descending coronary artery (LAD) and heart in the two positions differed individually, and were strongly related to the body mass index (BMI). Image fusion of the CT scans revealed that prone positioning was detrimental if the heart was situated distant from the chest wall in the supine position, but moved to the chest wall in the prone position. For characterization of the geography of the heart and the breast, the median distance between the LAD and the chest wall (d_{median}), and the heart area included in the radiation field on a single CT scan at the middle of the heart in the supine position (A_{heart}) proved most appropriate. *Conclusion.* A validated statistical model, utilizing the BMI, d_{median} and A_{heart}, permits individualized positioning for maximum heart protection.

Although breast irradiation after surgery for breast cancer contributes to an improved survival, it may cause radiation sequelae of the heart [1]. The most significant changes leading to radiation-induced heart disease occur in the cardiac micro- and macrovasculature [2]. In left-sided cases, the left anterior descending coronary artery (LAD) is often situated in the tangential fields and regarded as an organ at risk (OAR) [3,4].

Among other methods, individual positioning is a feasible approach through which to control the radiation dose to the OARs [5,6]. While prone positioning dramatically reduces the lung dose, reduction of the heart exposure is controversial [5–11]. Some studies indicated that prone positioning is preferable for heart protection in general [5], whereas others did not reveal a significant advantage of prone positioning over supine positioning in the overall population [7–10]. Three investigations demonstrated individual variability, but no specific patient-related feature was identified in favor of a particular treatment setup [6,9,10]. Others found an association between breast size and the benefit of prone positioning [8,11].

In this prospective study, the goal was the identification of patient-related parameters via which to predict the preferable positioning mode for breast radiotherapy in clinical practice.

Material and methods

The study was approved by the Institutional Review Board of the University of Szeged, and all the enrolled patients gave their written informed consent to participation. Consecutive patients with left-sided breast cancer requiring radiotherapy of the operated breast were included throughout the study.

CT-based three-dimensional (3D) treatment planning [XIO® (Elekta) vs. 4.2.0, convolution algorithm for photon dose calculation] was performed in both supine and prone positions, as detailed previously [7]; following the collection of dosimetric data,

Correspondence: Z. Kahán, University of Szeged, Korányi Fasor 12, 6720 Szeged, Hungary. Tel: +36 62 545406. Fax: +36 62 545922. E-mail: kahan.zsuzsanna@med.u-szeged.hu

(Received 10 January 2013; accepted 18 February 2013)

ISSN 0284-186X print/ISSN 1651-226X online © 2014 Informa Healthcare
DOI: 10.3109/0284186X.2013.781674

the treatment position chosen for radiotherapy was that which best spared the LAD and heart. The breast tissue was visualized on CT, and the clinical target volume was contoured at the chest wall/breast parenchyma interface, 4 mm from the skin, cranially the head of clavicle, medially the border of the sternum, laterally and caudally the visible breast parenchyma/connective tissue verge. 3D image reconstruction was used for checking delineation. Planning target volumes (PTVs) were generated by the addition of 3D 5-mm margins to the clinical target volume limited 4 mm from the skin. The heart and LAD were defined according to published recommendations [4,12] (Figure 1). Equivalent target and OAR volume contouring in either setup was ensured by one author (ZK). Treatment plans were developed by applying conventional 6-MV tangential photon fields set up isocentrically, and a median of 2 (1–3) individually weighted 6/15-MV segmental fields superimposed on the tangential fields by using a multileaf collimator. A mean dose to the PTV of 50 Gy was aimed at. For the analysis of dose distribution, the volume receiving 95–107% of the total dose ($V_{95–107\%}$), and the doses received by 5% and 95% of the PTV ($D_{5\%}/D_{95\%}$), the healthy tissue conformity index (HTCI) and the conformation number (CN) were calculated using the following equations: $\mathrm{HTCI} = \frac{TV_{RI}}{V_{RI}}$ and $\mathrm{CN} = \frac{TV_{RI}}{TV} \times \frac{TV_{RI}}{V_{RI}}$ (TV: Target volume, i.e. PTV; TV_{RI}: Target volume covered by the reference isodose; V_{RI}: Volume of the reference isodose) [13].

The radiation exposures of the OARs (the volume of the ipsilateral lung receiving >20 Gy [V_{20Gy}], the mean doses to the ipsilateral lung [MLD], LAD [MD_{LAD}] and heart, the volume of the LAD recieving >20 Gy [$V_{20GyLAD}$], the volume of the heart receiving >25 Gy [$V_{25Gyheart}$], and the volume of the contralateral breast receiving >5 Gy [V_{5Gy}]), were registered in both positions. The advantage of the prone over the supine position was analyzed in terms of the differences between MD_{LAD} and $V_{25Gy\text{-}heart}$ in the two positions (ΔMD_{LAD} and $\Delta V_{25Gyheart}$) in relation to the patient characteristics (the breast volume and the body mass index [BMI]).

In order to compare the geographies of the index breast, the heart and the LAD in the two positions, corresponding CT images acquired in the supine and prone positions were fused within the radiotherapy planning system. Perfect 3D image fusion was aimed at in the horizontal, sagittal and coronal planes. As the most important geographical pivots of breast radiotherapy planning, the sternum and anterior chest wall were selected to corroborate the adequacy of image matching. The image fusion indicated that the heart was situated at individual distances from the chest wall in the supine position, whereas in the prone position it almost always lay adjacent to the chest wall (Figure 2). The median of the shortest distances between the anterior surface of the LAD and the chest wall representing the middle of the LAD (d_{median}) and the area of the heart (measured with the DicomWorks vs. 1.3.5 software) in the radiation field on the same CT scan in the supine position (A_{heart}) typified the anatomical situation (Figure 1a insert).

Correlation-regression analysis, paired and independent sample t-tests, receiver operating characteristics (ROC) analyses, multivariate logistic regression and multiple linear regression models were used. The separability of the dataset was verified through artificial intelligence classification methods: Support Vector Machine (SVM) with dot kernel and Neural Net with one hidden layer [14,15]. Classification methods were trained on 83 cases and validated on a set of 55 further cases. A 1000-times random cross-validation method was applied to the overall dataset. Sensitivity and specificity were calculated with supine positioning as positive determinant in the model.

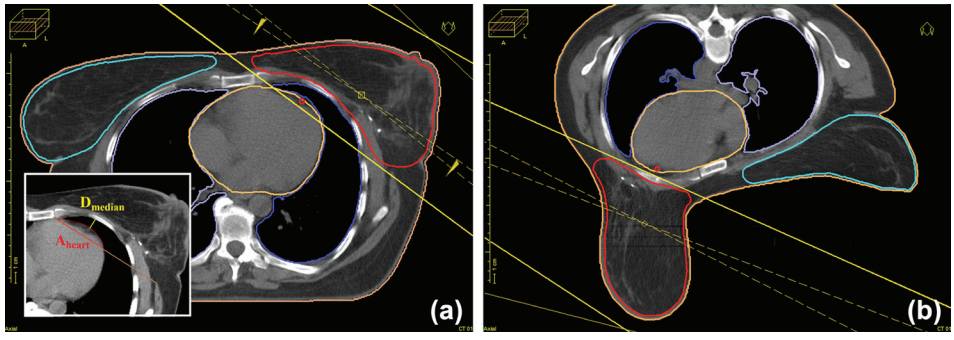


Figure 1. Typical PTV and OAR contouring and field setup in the supine (**a**) and prone (**b**) positions. The shortest distance between the anterior surface of the LAD and the chest wall (d_{median}) and the surface area of the heart in the radiation field (A_{heart}) are measured at the middle of the LAD in the supine position (insert).

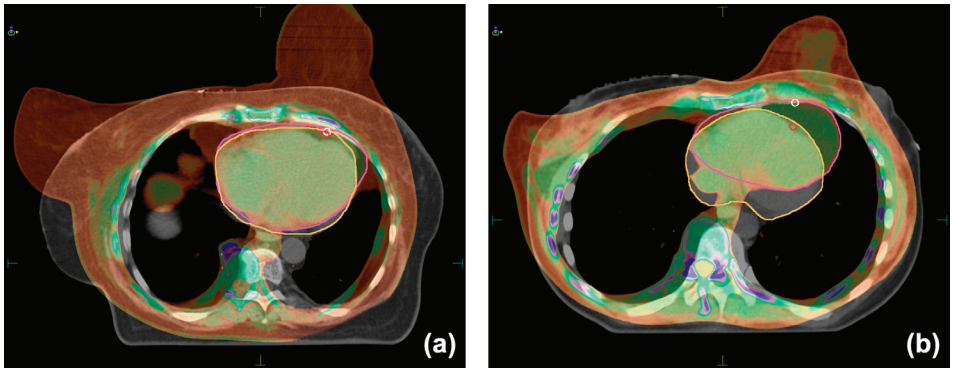


Figure 2. Image fusion of the CT scans acquired in the prone versus the supine position illustrates that, if the heart is adjacent to the chest wall in both positions, the prone position helps prevent heart exposure through separation of the breast from the chest wall (**a**); if the heart is distant from the chest wall in the supine position, the dose to the heart is increased due to its shift to the chest wall in the prone position (**b**). (greyscale: supine, color: prone).

Results

Heart and LAD doses

Table I shows the baseline characteristics of the two cohorts of patients included in the entire study; there was no significant difference between the two groups. In the population of the first 83 patients (training set), the mean values of the radiation doses to the ipsilateral lung, LAD and heart were significantly lower in the prone than in the supine position (Table II); the dose homogeneity was worse, but the HTCI and CN values indicated that less healthy tissue was exposed to radiation in the prone position (Table III). ΔMD_{LAD} and $\Delta V_{25Gyheart}$ displayed individual differences: in about half of the patients, the radiation dose to the heart was lower in the prone position; in the others, it was higher or not significantly different (Figure 3). In 40% and 19% of the patients, respectively, MD_{LAD} and $V_{25Gyheart}$ were significantly higher in the prone position. The doses to the LAD and the heart were both lower in the supine position in 22 cases, and both lower in the prone position in 45 cases, while in 16 cases the results were discordant: from the aspect of the LAD dose, 14 cases favored supine positioning, and two cases prone positioning.

The effect of the BMI

The difference in heart or LAD doses in the prone versus supine position did not differ according to whether the volume of the operated breast was $<$ or $\geq$ the median value of 900 cm^3 (Figure 3). In ROC analyses and multivariate analyses including the BMI, the PTV and the heart volume, the BMI gave the best results for ΔMD_{LAD} and $\Delta V_{25Gyheart}$ (Table IV); ROC curve analysis indicated that patients with BMI ≥ 26.3 kg/m^2 benefited most from prone positioning. In the multivariate logistic regression models relating to the benefit of prone positioning, only the BMI remained significant. The likelihood of a one unit increase in ΔMD_{LAD} and $\Delta V_{25Gyheart}$ for prone versus supine positioning was increased, with OR = 1.256 (1.103–1.430, p = 0.001) and OR = 1.404 (95% CI 1.180–1.672, $p < 0.001$), respectively, for every 1 kg/m^2 increase in BMI.

The effect of the anatomical variables

Since the volumes of the breast ($R = 0.681$, $p < 0.001$) and heart ($R = 0.440$, $p < 0.001$) correlated significantly with the BMI, we postulated that the BMI is a resultant of the variables that determine the position and size of the target volume and the OARs, and

Table I. Baseline characteristics of the patients included in the study (n = 138).

	Training set (n = 83)			Validation set (n = 55)		
Variable	Median	Mean ± SE	Range	Median	Mean ± SE	Range
Age (years)	60.2	59.4 ± 1.1	31.1–79.1	59.1	58.4 ± 1.4	26.0–76.5
Weight (kg)	73.0	73.4 ± 1.4	46.0–112.0	75.0	77.1 ± 1.9	49.0–120.0
Height (cm)	162.0	162.4 ± 0.7	149.0–178.0	162.0	161.1 ± 0.8	140.0–179.0
BMI (kg/m^2)	27.5	27.8 ± 0.5	17.1–38.9	28.4	29.7 ± 0.7	20.3–44.1
Breast volume (cm^3)	897.0	983.1 ± 46.8	197.0–2448.0	1061.0	1050.6 ± 65.1	257.0–2838.0
Heart volume (cm^3)	515.0	522.0 ± 11.6	307.0–965.0	540.0	553.5 ± 14.7	360.0–862.0
d_{median} (cm)	1.3	1.33 ± 0.1	0.4–2.2	1.4	1.39 ± 0.1	0.3–3.3
A_{heart} (mm^2)	549.0	599.9 ± 43.9	0–1820.0	455.0	476.9 ± 50.0	0–1627.0

Table II. Radiation doses to the OARs. Mean values ± SE are shown.

	LAD		Heart		Ipsilateral lung		Contralateral breast
n = 83	Mean dose (Gy)	V_{20Gy} (%)	Mean dose (Gy)	V_{25Gy} (%)	MLD (Gy)	V_{20Gy} (%)	V_{5Gy} (%)
Prone	11.06 ± 0.79	21.91 ± 2.18	2.18 ± 0.15	2.01 ± 0.25	0.99 ± 0.18	1.33 ± 0.23	1.57 ± 0.25
Supine	13.70 ± 0.79	29.26 ± 1.98	2.89 ± 0.19	3.54 ± 0.37	6.29 ± 0.29	11.87 ± 0.61	1.07 ± 0.32
p*	0.014	0.010	0.001	< 0.001	< 0.001	< 0.001	0.154

*paired t-test

their relationship. We therefore studied the situation of the operated breast and the heart via image fusion of the prone and supine CT scans. When BMI ≥ 26.3 kg/m^2, the heart usually lay adjacent to the chest wall in both positions, and removal of the breast from the chest wall to the hanging position involved a radiation field geometry change and a heart dose reduction (Figure 2a). When BMI < 26.3 kg/m^2, the heart was situated distant from the chest wall, but fell many centimeters anterior in the prone position, usually directly to the chest wall, which favored an increased heart exposure, despite the separation of the breast from the chest wall (Figure 2b). For quantitative characterization, d_{median} and A_{heart} were selected (Figure 1a insert). d_{median} proved to correlate inversely with $\Delta V_{25Gyheart}$ ($R = -0.477$, $p < 0.001$), ΔMD_{LAD} ($R = -0.567$, $p < 0.001$), A_{heart} ($R = -0.424$, $p < 0.001$) and the BMI ($R = -0.344$, $p < 0.001$).

Predictive model

None of these predictors alone resulted in a good classification, and linear regression models were therefore developed from these anatomical features and the BMI for the prediction of optimal positioning, with ΔMD_{LAD} and $\Delta V_{25Gyheart}$ as dependent variables. A good classification of the cases (depending on the choice of cut-off point) was achieved, based on the estimated dose differences. With a single cut-off point, a case was classified to prone positioning when the predicted value exceeded the cut-off point. Based on the data of the first 83 cases, reasonable sensitivity and specificity values were verified (Table V).

The correctness of the model was checked in various ways: 1) In a validation set of 55 cases, the sensitivity and specificity were 72.2% and 89.2% for ΔMD_{LAD}, and 90.2% and 35.7% for $\Delta V_{25Gyheart}$, respectively; 2) In a 1000-times random cross-validation procedure on the dataset from the 138 cases, both the proportion of misclassified patients and the extent of misclassification were assessed via the dose difference (Table V): for ΔMD_{LAD} a threshold of 0.9 Gy, and for $\Delta V_{25Gyheart}$ a cut-off point of 0.75% seemed preferable; 3) For the overall dataset with the SVM artificial intelligence model, the sensitivity and specificity were 81.0% and 87.0% (ΔMD_{LAD}) and 88.1% and 55.3% ($\Delta V_{25Gyheart}$), respectively; with Neural Net they were 85.7% and 76.0% (ΔMD_{LAD}) and 88.0% and 60.5% ($\Delta V_{25Gyheart}$), respectively.

Discussion

We have demonstrated that, although prone positioning dramatically reduces the dose to the lung, it increases the LAD and heart doses in a significant proportion of the cases. The geometries of the heart, chest wall and breast are the most important determinants of the cardiac dose. For consideration of the patient's anatomic variables, a supine CT scan seems preferable. The BMI additionally facilitates the choice of treatment setup via rough orientation, or its use in a predictive model. As a tool for individualized heart protection in clinical practice, the model based on BMI, d_{median} and A_{heart} ensures good sensitivity and specificity.

For calculation of the heart dose in the prone versus supine treatment setup, different approaches have been implemented. In 51 IMRT cases [5,16], and in a series of 200 patients treated with conformal radiotherapy [6,17], the in-field heart volume served as a surrogate marker of radiation heart exposure. Other investigations have used conformal radiotherapy and either V_{20Gy} [9] or V_{30Gy} [9,10]. In their pioneering study of the LAD and heart doses, Kirby et al.

Table III. Dose homogeneity and conformity according to the treatment position ($p < 0.001$ in all cases).

	$V_{95\%-107\%}$ (%, mean ± SE)	$D_{5\%}/D_{95\%}$ (mean ± SE)	Healthy tissue conformation index (HTCI) (mean ± SE)	Conformation number (CN) (mean ± SE)
Supine	91.30 ± 7.48	1.10 ± 0.09	0.80 ± 0.06	0.58 ± 0.05
Prone	88.34 ± 7.24	1.14 ± 0.09	0.940 ± 0.08	0.682 ± 0.06

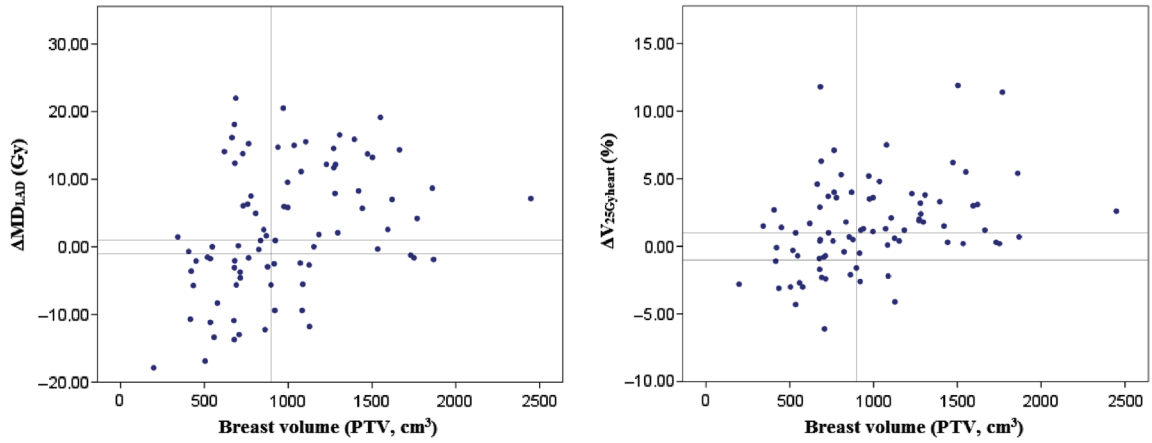


Figure 3. Benefit of the prone position (the difference in LAD and heart doses between the prone and supine treatment positions) in the overall population of 83 patients as a function of breast size (LAD: ΔMD_{LAD}, heart: $\Delta V_{25Gyheart}$).The vertical line indicates the median PTV value of 900 cm^3, the horizontal lines indicate the dose differences regarded as clinically not significant, ± 1 Gy (MD_{LAD}) and ± 1% ($V_{25Gyheart}$).

[8] analyzed the mean heart dose, whereas we preferred $V_{25Gyheart}$ for the assessment of the high-dose volume of the heart. In fact, this measure was shown to correlate with increased cardiac mortality [18,19], and has been used as an indicator of heart dose in more recent studies too [20,21]. For the LAD, because of its very small volume, we considered the MD_{LAD} appropriate. Our results on the LAD dose in 138 cases are comparable with those of Kirby et al. [8].

The proportion of patients with lower heart exposure in the prone than in the supine position varies. As concerns the in-field heart volume, Lymberis et al. found that 46 (87%) of 53 left breast cancer patients benefited from prone positioning, while the supine position was preferable in only five cases (10%) [11]. Kirby et al. [8] concluded that prone positioning reduced the heart and LAD doses in about two-thirds of the patients, but was detrimental in approximately one-third of the cases, as regards the mean doses to both the LAD and the heart. Formenti et al. [6] demonstrated that the in-field volume of the heart was reduced in the prone position in 85% of the cases. Relative to all of the reported studies, we found a higher proportion of patients who benefited from supine positioning: in 40% and 19% of the patients, the dose to the LAD or heart, respectively, was significantly lower in the supine than in the prone position. The underlying causes of this difference might include the methods and indicators we used. The relatively high number of our cases with a preferable LAD dose in the supine position draws attention to the need for a refined comprehensive OAR protection approach. The strength of our study is that it provided robust dose-volume data on both the heart and the LAD. The reliability of the findings is supported by the use of justified methods (equivalent volumes of target and OARs in the two positions), and the consistency regarding the sensitivity and specificity of the model.

The question of which OAR is most important as concerns radiation-induced heart damage remains unanswered. The clinical results indicate that the high-dose volume of the heart is the key determinant of the long-term outcome, but the apparently logical role of the LAD has not been completely clarified [1–4]. The situation is further complicated by the finding that those who gain from prone positioning through heart protection are not always those who benefit through LAD exposure. The proportion of

Table IV. ROC curve analyses of selected patient characteristics (n = 83) as predictors of the benefit of prone positioning for heart protection during breast radiotherapy.

	AUC MD_{LAD} ± SE	AUC $V_{25Gyheart}$ ± SE
Weight	0.692 ± 0.061 (p = 0.003)	0.793 ± 0.056 (p < 0.001)
Height	0.370 ± 0.062 (p = 0.043)	0.417 ± 0.068 (p = 0.238)
BMI	0.740 ± 0.056 (p < 0.001)	0.825 ± 0.054 (p < 0.001)
Breast volume (PTV)	0.722 ± 0.058 (p = 0.001)	0.813 ± 0.047 (p < 0.001)
Heart volume	0.652 ± 0.062 (p = 0.018)	0.661 ± 0.064 (p = 0.022)
d_{median}	0.785 ± 0.050 (p < 0.001)	0.730 ± 0.060 (p = 0.001)
A_{heart}	0.868 ± 0.039 (p < 0.001)	0.852 ± 0.050 (p < 0.001)

Table V. Classification measures for ΔMD_{LAD} and $\Delta V_{25Gyheart}$ using a single discrimination threshold.

	Cut-off point (Gy)	Sensitivity (%)	Specificity (%)	Extent of wrong estimation, decision: prone (Gy, mean ± SD)	Extent of wrong estimation, decision: supine (Gy, mean ± SD)
ΔMD_{LAD} (Gy)	−0.6	66.6	91.1	2.5 ± 3.9	−0.7 ± 1.0
	−0.3	70.8	90.7	2.6 ± 3.6	−0.8 ± 1.1
	0	74.4	90.0	2.4 ± 3.4	−0.9 ± 1.3
	0.3	77.7	88.9	2.1 ± 3.0	−1.2 ± 1.6
	0.6	80.7	87.5	1.7 ± 2.6	−1.7 ± 1.9
	0.9	83.4	86.0	1.5 ± 2.4	−2.0 ± 2.2
	1.2	85.4	83.6	1.1 ± 2.3	−2.3 ± 2.8
	1.5	86.5	81.7	1.1 ± 2.2	−3.0 ± 3.7
	1.8	86.8	79.9	1.3 ± 2.3	−3.5 ± 4.2
$\Delta V_{25Gyheart}$ (%)	0	47.9	89.7	1.19 ± 1.43	−0.39 ± 0.47
	0.25	56.2	88.8	1.14 ± 1.40	−0.47 ± 0.52
	0.50	63.2	85.9	1.05 ± 1.40	−0.52 ± 0.52
	0.75	72.4	82.4	0.86 ± 1.37	−0.54 ± 0.82
	1	78.8	77.7	0.82 ± 1.43	−0.65 ± 0.94
	1.25	84.0	74.0	0.75 ± 1.47	−0.75 ± 0.96
	1.50	87.4	77.0	0.71 ± 1.51	−0.98 ± 1.05
	1.75	89.9	62.1	0.63 ± 1.17	−1.14 ± 1.24

our cases with such discordant results was 19%; it is noteworthy that ΔMD_{LAD} was significantly smaller in these cases than in the concordant cases (data not shown). A similar disagreement of heart and LAD data was described by Kirby et al. [8], and Aznar et al. [22] reported that 8/24 patients irradiated in the supine position received high LAD doses despite the heart doses not exceeding the dose constraint.

We consider that the LAD dose is of prime interest: 1) Irradiation of the LAD has more significant consequences than irradiation of a small part of the myocardium [1,2]; 2) The LAD, situated on the anterior surface of the myocardium, closest to the radiation beam, may be regarded as a surrogate indicator of the radiation harm. Further, the systematic displacement of the heart by the prone setup is greatest at the supero-lateral aspect of the heart [23], where the LAD runs. Kirby et al. [8] observed that only patients in whom the breast tissue is pulled anteriorly are likely to gain from prone treatment. In contrast, we found that only patients whose heart lay adjacent to the chest wall in both positions were likely to gain from the prone position, because the separation of the breast from the chest wall could prevent heart irradiation in these cases. Our strategy in individual cases is to consider the LAD dose first, then $V_{25Gyheart}$, and finally the lung dose. Since the acquisition of two sets of CT to compare the dosimetry in the two positions contravenes radiation-hygienic and economic principles, we are aware of the need for a simple method for routine practice. In fact, different approaches have been reported as clinical tools for prediction of the preferable patient setup. Kirby et al. [8] concluded that prone positioning may be detrimental in left-sided breast cancer patients with small breasts. In a group of 198 patients where no valid classifier, including the breast size, was identified, Zhao et al. [17] developed a two-step decision-analysis algorithm using a weighted SVM. Based on the anatomical features detected on a prone CT series, this classified patients to prone radiotherapy or to a second CT in the supine position for comparison. Their strategy is in contrast with ours: with CT in all cases in the prone position first, the strongest determinant of the benefit in the supine setup, i.e. the distance between the heart and the chest wall, is missed. We identified three relevant anatomical features that characterize the size and geography of the index breast and the OARs, and developed a stable statistical model that we validated in 138 cases. With the aim of easy acquisition of these decisive data (and the avoidance of two CT series), we are currently testing a simple clinical method. A single CT slice image representing the middle of the heart is acquired by using the AP scout view in the supine position for the selection of the correct transversal plane. On that CT scan, d_{median} and A_{heart} are measured after placing a straight line between the back muscle and the lateral edge of the sternum, representing the posterior edge of the radiation fields; the initial experience as concerns finding the correct slice, and using the data coming from it, is promising.

Interestingly, in a prospective study of 108 patients with negative pre-radiotherapy myocardial perfusion scans, an independent risk factor of developing perfusion defects during the follow-up period was found a BMI ≥ 25 kg/m^2, probably involving larger doses to the critical structures of the heart during radiotherapy in the supine position [24]. In a recent analysis

of 100 radiotherapy plans in the supine position, among different patient-related features, the BMI was found the only independent predictor of 'suboptimal heart anatomy' defined as a $V_{25Gy\ heart} \geq 6\%$ involving increased risk of radiation-induced coronary artery disease [20]. Although the BMI has been found useful for rough orientation in our practice, the threshold of 26.3 kg/m^2 may not be most appropriate for the selection of those who benefit most from prone positioning in other populations.

In summary, the main outcomes of our study are a clarification of the interplay of various patient-related parameters that influence the exposure of the different structures of the heart during breast radiotherapy in the prone and supine positions, and the development of a practical tool for prediction of the preferable treatment setup in individual patients. Naturally, our results need independent validation.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

[1] McGale P, Darby SC, Hall P, Adolfsson J, Bengtsson NO, Bennet AM, et al. Incidence of heart disease in 35,000 women treated with radiotherapy for breast cancer in Denmark and Sweden. Radiother Oncol 2011;100:167–75.

[2] Andratschke N, Maurer J, Molls M, Trott KR. Late radiation-induced heart disease after radiotherapy. Clinical importance, radiobiological mechanisms and strategies of prevention. Radiother Oncol 2011;100:160–6.

[3] Nilsson G, Holmberg L, Garmo H, Duvernoy O, Sjögren I, Lagerqvist B, et al. Distribution of coronary artery stenosis after radiation for breast cancer. J Clin Oncol 2012;30: 380–6.

[4] Feng M, Moran JM, Koelling T, Chughtai A, Chan JL, Freedman BL, et al. Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. Int J Radiat Oncol Biol Phys 2011;79:10–8.

[5] Formenti SC, Gidea-Addeo D, Goldberg JD, Roses DF, Guth A, Rosenstein BS, et al. Phase I–II trial of prone accelerated intensity modulated radiation therapy to the breast to optimally spare normal tissue. J Clin Oncol 2007;25:2236–42.

[6] Formenti SC, DeWyngaert JK, Jozsef G, Goldberg JD. Prone vs supine positioning for breast cancer radiotherapy. JAMA 2012;308:861–3.

[7] Varga Z, Hideghéty K, Mezö T, Nikolényi A, Thurzó L, Kahán Z. Individual positioning: A comparative study of adjuvant breast radiotherapy in the prone versus supine position. Int J Radiat Oncol Biol Phys 2009;75:94–100.

[8] Kirby AM, Evans PM, Donovan EM, Convery HM, Haviland JS, Yarnold JR. Prone versus supine positioning for whole and partial-breast radiotherapy: A comparison of non-target tissue dosimetry. Radiother Oncol 2010;96:178–84.

[9] Griem KL, Fetherston P, Kuznetsova M, Foster GS, Shott S, Chu J, et al. Three-dimensional photon dosimetry: A comparison of treatment of the intact breast in the supine and prone position. Int J Radiat Oncol Biol Phys 2003; 57:891–9.

[10] Buijsen J, Jager JJ, Bovendeerd J, Voncken R, Borger JH, Boersma LJ, et al. Prone breast irradiation for pendulous breasts. Radiother Oncol 2007;82:337–40.

[11] Lymberis SC, Dewyngaert JK, Parhar P, Chhabra AM, Fenton-Kerimian M, Chang J, et al. Prospective assessment of optimal individual position (prone versus supine) for breast radiotherapy: Volumetric and dosimetric correlations in 100 patients. Int J Radiat Oncol Biol Phys 2012; 84:902–9.

[12] Taylor CW, Povall JM, McGale P, Nisbet A, Dodwell D, Smith JT, et al. Cardiac dose from tangential breast cancer radiotherapy in the year 2006. Int J Radiat Oncol Biol Phys 2008;72:501–7.

[13] Feuvret L, Noël G, Mazeron JJ, Bey P. Conformity index: A review. Int J Radiat Oncol Biol Phys 2006;64:333–42.

[14] Chang RF, Wu WJ, Moon WK, Chou YH, Chen DR. Support vector machines for diagnosis of breast tumors on US images. Acad Radiol 2003;10:189–97.

[15] Bishop CM. Neural networks. In: Bishop CM, editor. Pattern recognition and machine working, 1st ed. New York: Springer; 2006. p. 225–90.

[16] DeWyngaert JK, Jozsef G, Mitchell J, Rosenstein B, Formenti SC. Accelerated intensity-modulated radiotherapy to breast in prone position: Dosimetric results. Int J Radiat Oncol Biol Phys 2007;68:1251–9.

[17] Zhao X, Wong EK, Wang Y, Lymberis S, Wen B, Formenti S, et al. A support vector machine (SVM) for predicting preferred treatment position in radiotherapy of patients with breast cancer. Med Phys 2010;37:5341–50.

[18] Gyenes G, Gagliardi G, Lax I, Fornander T, Rutqvist LE. Evaluation of irradiated heart volumes in stage I breast cancer patients treated with postoperative adjuvant radiotherapy. J Clin Oncol 1997;15:1348–53.

[19] Gyenes G, Rutqvist LE, Liedberg A, Fornander T. Long-term cardiac morbidity and mortality in a randomized trial of pre- and postoperative radiation therapy versus surgery alone in primary breast cancer. Radiother Oncol 1998;48:185–90.

[20] Evans SB, Sioshansi S, Moran MS, Hiatt J, Price LL, Wazer DE. Prevalence of poor cardiac anatomy in carcinoma of the breast treated with whole-breast radiotherapy: Reconciling modern cardiac dosimetry with cardiac mortality data. Am J Clin Oncol 2012;35:587–92.

[21] Hjelstuen MH, Mjaaland I, Vikström J, Dybvik KI. Radiation during deep inspiration allows loco-regional treatment of left breast and axillary-, supraclavicular- and internal mammary lymph nodes without compromising target coverage or dose restrictions to organs at risk. Acta Oncol 2012;51:333–44.

[22] Aznar MC, Korreman SS, Pedersen AN, Persson GF, Josipovic M, Specht L. Evaluation of dose to cardiac structures during breast irradiation. Br J Radiol 2011;84: 743–6.

[23] Chino JP, Marks LB. Prone positioning causes the heart to be displaced anteriorly within the thorax: Implications for breast cancer treatment. Int J Radiat Oncol Biol Phys 2008;70:916–20.

[24] Evans ES, Prosnitz RG, Yu X, Zhou SM, Hollis DR, Wong TZ, et al. Impact of patient-specific factors, irradiated left ventricular volume, and treatment set-up errors on the development of myocardial perfusion defects after radiation therapy for left-sided breast cancer. Int J Radiat Oncol Biol Phys 2006;66:1125–34.

III. Kahán, Z., **Rárosi, F.**, Gaál, S., Cserháti, A., Boda, K., Darázs, B., Kószó, R., Lakosi ,F., Gulybán, Á., and Varga, Z. (2018). A simple clinical method for predicting the benefit of prone vs. supine positioning in reducing heart exposure during left breast radiotherapy. Radiotherapy and Oncology 126(3), 487–492. IF: 4.942

Radiotherapy and Oncology 126 (2018) 487–492

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Breast radiotherapy

A simple clinical method for predicting the benefit of prone *vs.* supine positioning in reducing heart exposure during left breast radiotherapy

Zsuzsanna Kahán [a,*], Ferenc Rárosi [b], Szilvia Gaál [a], Adrienn Cserháti [a], Krisztina Boda [b], Barbara Darázs [a], Renáta Kószó [a], Ferenc Lakosi [c,d], Ákos Gulyban [c,e], Philippe A. Coucke [c], Zoltán Varga [a]

[a] *Department of Oncotherapy;* [b] *Department of Medical Informatics, University of Szeged, Hungary;* [c] *Department of Radiation Oncology, University Hospital of Liège, Belgium;* [d] *Institute of Diagnostic Imaging and Radiation Oncology, Health Center, Kaposvár University, Hungary;* [e] *Radiation Oncology Department, Europe Hospitals Brussels, Belgium*

ARTICLE INFO

Article history:
Received 26 April 2017
Received in revised form 20 December 2017
Accepted 22 December 2017
Available online 17 January 2018

Keywords:
Breast radiotherapy
Clinical tool
Heart protection
Individual positioning
LAD protection

ABSTRACT

Background and purpose: The benefit of reduced radiation heart exposure in the prone *vs.* supine position individually differs. In this prospective cohort study, the goal was to develop a simple method for the operation of a validated model for the prediction of preferable treatment position during left breast radiotherapy.

Material and methods: In 100 cases, a single CT slice was utilized for the collection of the needed patient-specific data (in addition to body mass index, the distance of the LAD from the chest wall and the area of the heart included in the radiation fields at the middle of the heart in the supine position). Outcome was analyzed in relation to the full CT series acquired in both positions and dosimetric data.

Results: Great consistency was found between the tested and original method regarding sensitivity and specificity. The prioritization of LAD dose, and the use of heart dose and position-specific dose constraints as safety measures ensure sensitivity and specificity values of 82.8% and 87.3%, respectively. In an additional "routine clinical practice" series of 60 patients the new method seemed feasible in routine clinical practice. External testing on a 28-case series indicated similar accuracy.

Conclusion: We consider this simple clinical tool appropriate for assisting individual positioning aiming at maximum heart protection during left breast irradiation.

© 2018 Elsevier B.V. All rights reserved. Radiotherapy and Oncology 126 (2018) 487–492

Radiotherapy is an essential component of the management of early breast cancer. The outcome in most cases is favorable, the majority of the affected patients become long survivors. Breast radiotherapy, however, may increase the risk of non-breast cancer-related morbidities, among which heart diseases rank the first [1,2]. Radiation-induced heart damage clearly depends on the dose exposed to its different structures (3,4). While older radiotherapy practices caused more significant late hazards, heightened awareness and the use of current technical developments make this danger much lower [1,4,5]. Although the application of modern radiotherapy planning and delivery significantly improves the control of radiation dose, in many cases a part of the heart, and especially the left anterior descending artery (LAD) located to its anterior surface still receive a dose sufficient to cause long-term adverse effects. Radiogenic diffuse myocardium damage including microvasculature abnormalities, degenerative cardiomyocyte and interstitial fibrotic changes may be controlled if not extensive, but the damage of the macrovasculature indistinguishable from coronary arteriosclerosis due to other causes more likely lead to a fatal outcome [3,6–8]. The exposure of the heart and the LAD are related [9–11], and irradiation-related cardiac morbidity and mortality are considered to be consequences of late manifesting coronary artery damage. Hence the verification and control of the dose to the LAD, is of prime importance [8,9,11,12].

With the aim of cardiac dose sparing and avoidance, numerous new methods have been developed [4,5]. These include the breath-holding techniques, prone positioning (both operate by separating the heart and the radiation fields), IMRT, proton irradiation or the reduction in the volume to be irradiated, partial breast irradiation (PBI). A significant increase in the number of clinical studies [11–20], and a recent survey on clinical practice [21] suggest that prone positioning has become an alternative of conventional supine positioning in some centers. Prone positioning always provides dramatic reduction in the ipsilateral lung dose, and in many cases significantly reduces heart exposure, too. A potential disadvantage is inferior repositioning accuracy, which may be improved with experience [18] or may be compensated by online daily correction [12,22].

* Corresponding author at: Department of Oncotherapy, University of Szeged, Korányi fasor 12, H-6720 Szeged, Hungary.
E-mail address: kahan.zsuzsanna@med.u-szeged.hu (Z. Kahán).

https://doi.org/10.1016/j.radonc.2017.12.021
0167-8140/© 2018 Elsevier B.V. All rights reserved.

Prone positioning was first invented for the irradiation of large-breasted women [23,24]. Indeed, since gravity pulls the breast away from the chest wall, the geometry of a pendulous breast and the tangential irradiation fields gets advantageous in the prone position [12]. Taking the overall population of breast cancer patients, however, prone positioning has such effect in 77–87% of cases only [11,14,15,19]. As a consequence, the position-dependent dose to the LAD or heart also individually differs [11,19,20]. Different approaches exist for selecting the optimal position in left breast cancer cases. Kirby et al. found that a PTV > 1000 cm^3 favors prone positioning [11]. Zhao et al. developed a two-step decision-analysis algorithm that, based on the anatomical features detected on a prone CT series, classified patients to prone radiotherapy or to a second CT in the supine position for comparison [25]. We have demonstrated that a statistical model utilizing 3 anatomical determinants (the body mass index [BMI], the distance of the LAD from the chest wall and the area of the heart included in the radiation fields at the middle of the heart in the supine position) of the patient gives accurate estimates on the benefit of one specific position over the other by means of LAD or heart doses [19]. Here we report on an original method for providing the necessary patient-specific data based on a single CT slice image representing the middle of the heart. In this prospective study, following the validation of the clinical tool, also its routine use has been tested on a separate series of cases.

Patients and methods

The study was approved by the Institutional Review Board of the University of Szeged, and all the enrolled patients gave their written informed consent to participation. Eligible patients needed postoperative left breast radiotherapy.

Outline of the study

First, a single CT slice image at the middle of the heart (reference plane, P_{ref}) was acquired with the help of an AP scout view in the supine position (Fig. 1A). On that CT scan, the shortest distance between the anterior surface of the LAD and the chest wall (D_{med}) and the area of the heart (A_{heart}) included in the radiation fields were measured after placing a straight line between the border of the ipsilateral latissimus dorsi muscle and the lateral edge of the sternum (Fig. 1B); these data (representing the topography of the heart) were introduced to the calculator together with the patient's BMI (which correlated with the volumes of the breast and heart) as previously described in detail [19]. The calculator based on a validated statistical model provided the estimated LAD and heart dose differences in the prone *vs.* supine position of the individual patient. In the first validation set of 100 patients, CT series were acquired in both the supine position and prone position. Conformal radiation treatment plans were generated in both positions using conventional 6 MV tangential photon fields set up isocentrically and median 2 (1–3) individually weighted 6/15 MV segmental fields superimposed on the tangential fields using a multileaf collimator as described [18,19]. Wedges were used in almost all supine radiation plans. A mean dose to the PTV of 50 Gy (25 fractions) and a uniform distribution (−5% + 7%) of the prescribed dose to 95% of the PTV, were aimed at. The consistency of all contouring activities had been ensured by a chief radiation oncologist (ZK) and an experienced radiologist (AC) [26]. Equivalent heart and LAD volume contouring in either setup was ensured by one author (ZK). In the next "routine clinical practice" set of 60 patients, the acquisition of a single series of CT images according to the suggestion of the calculator was aimed at, and a second CT series was taken only if any of the dose constraints approved for the specific position were not reached in the position suggested by the calculator. In this series of patients' dose constraints were specified on the basis of previously recorded data. The upper range limits of the 90% percentile of dosimetry data in the preferred position were the following: mean LAD dose [MD_{LAD}]: 12.9 Gy and 12.5 Gy, $V_{25Gyheart}$: 2.4% and 4.7%, in the prone position and supine position, respectively. In true discordant cases, our strategy for selecting treatment position was to consider the LAD dose as a primary decisive factor.

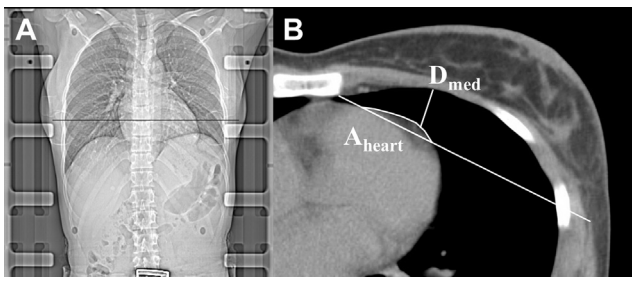


Fig. 1. The simple clinical tool generates patient-specific data to predict the benefit of prone positioning. After selecting the reference plane (P_{ref}) at the middle of the heart on the AP scout view (A), a single CT slice is acquired for the measurement of those determinants (D_{med} and A_{heart}) (B) which operate the calculator to provide estimates of the doses to the LAD or heart.

In the validation set, data on LAD and heart dose differences between the two treatment positions were extracted from the planning system and estimated by the calculator, whereas in the "routine clinical practice" series only the estimated dose differences were available. Analyses were performed on 1. the equivalence of the P_{ref} with the median plane of the full series of CT scans acquired in the supine position (P_{med}) and 2. the effect of plane miss on the patient-related determinants and choice of preferable position. The sensitivity and specificity of this simple clinical method were evaluated based on the dosimetry data obtained using the topogram for selecting the position (n = 100). In the "routine clinical practice" series, the acceptability of the position as predicted by the calculator, the LAD and heart doses achieved without taking 2 CT series, and the need of performing a second CT series and changing position or irradiation technique were analyzed.

External testing

The supine and prone CT series and supine topogram of patients included in the study "Individualized positioning for maximum heart and index breast protection during breast irradiation: comparative study between Prone and Supine (Approval: 26/09/2013, B707201318246) were retrospectively used for independent testing. The protocol of patient positioning, delineation and radiation treatment planning has been described [27].

First, P_{ref} was selected on the topogram. Then, the predictors BMI, D_{med}, A_{heart} as measured in P_{ref} were introduced to the calculator. As a second step, D_{med}, A_{heart} were also measured in $P_{med.}$ LAD and heart dose differences between the two treatment positions extracted from the planning system and estimated by the calculator were analyzed. Finally, the correctness of P_{ref} was evaluated.

Statistical methods

The calculator had been developed based on linear regression models utilizing the patients' anatomical features, with ΔMD_{LAD} and $\Delta V_{25Gyheart}$ as dependent variables [19]. With a single cut-off point, a case was classified to prone positioning when the predicted value exceeded that value. Thresholds were optimized based on sensitivity and specificity as calculated from previous

Table 1
Classification measures for ΔMD_{LAD} and $\Delta V_{25Gyheart}$ using a single discrimination threshold. Great consistency is seen between the original cohort [19] and the present series.

		Original method (double CT method, $n = 83$)		Simple tool (single CT method, $n = 100$)	
	Cut-off point	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
ΔMD_{LAD} (Gy)	−0.6	66.6	91.1	72.4	91.5
	−0.3	70.8	90.7	75.9	91.5
	0	74.4	90.0	75.9	91.5
	0.3	77.7	88.9	79.3	88.7
	0.6	80.7	87.5	82.8	87.3
	0.9	83.4	86.0	82.8	83.1
	1.2	85.4	83.6	86.2	81.7
	1.5	86.5	81.7	86.2	77.5
	1.8	86.8	79.9	93.1	76.1
$\Delta V_{25Gyheart}$ (%)	0	47.9	89.7	50	90.8
	0.25	56.2	88.8	58.3	89.5
	0.50	63.2	85.9	64	88
	0.75	72.4	82.4	68	85.3
	1	78.8	77.7	80	85.3
	1.25	84.0	74.0	84	81.3
	1.50	87.4	77.0	92	78.6
	1.75	89.9	62.1	96	74.6

Table 2
D_{med} and A_{heart} values (mean ± SD) as measured on P_{ref} *vs.* P_{med} in all cases or in correctly and incorrectly specified P_{ref} cases; the measurements were performed on 2 CT scans at the middle of the heart either identified with the help of an A-P scout view (P_{ref}) or selected from a full CT series (P_{med}).

	All cases ($n = 100$)		Correct plane ($n = 55$)		Plane miss ($n = 45$)	
	P_{ref}	P_{med}	P_{ref}	P_{med}	P_{ref}	P_{med}
D_{median} (cm)	1.27 ± 0.59	1.25 ± 0.67	1.35 ± 0.55	1.17 ± 0.63	1.18 ± 0.63	1.34 ± 0.71
A_{heart} (mm^2)	768.8 ± 487.4	671.6 ± 450.1	730.7 ± 537.4	721.5 ± 511.2	815.4 ± 419.5	610.5 ± 358.1

[19] and present data (Table 1). Sensitivity and specificity were calculated with supine positioning as positive determinant in the model. For ΔMD_{LAD} a threshold of 0.6 Gy, and for $\Delta V_{25Gyheart}$ a cut-off point of 1.0% were chosen. In the definition of the cut-off points, a sensitivity of 80% at the minimum and the maximum achievable value of specificity was required.

LAD and heart dose constraints achievable by selecting the preferable position were specified by percentage estimation. Statistical analysis was performed with SPSS 22.0 for Windows.

Results

Validation set

In 55/100 cases, P_{ref} was the same as P_{med} while in 28 and 17 cases, P_{ref} and P_{med} differed by 1 or more planes, respectively. More among the incorrectly defined P_{ref} cases were shifted toward the caudal than the cranial direction. This resulted in smaller mean D_{med} and larger mean A_{heart} values among the plane miss cases overall (Table 2). Within the whole series, no change in the frequency of plane misses could be detected by time. Incongruency among ΔMD_{LAD} and $\Delta V_{25Gyheart}$ in the supine and prone position as predicted by the calculator on the basis of P_{ref} *vs.* P_{med} data, was present in 14 and 18 of the cases, respectively; these were all of small numerical values (Fig. 2A, B). When the LAD and heart dose differences predicted by the calculator based on the P_{ref} values were compared with the original dosimetric data from plans generated in both positions, the suggestion proved invalid in 14 (MD_{LAD}) and 16 ($V_{25Gyheart}$) cases (Fig. 2C, D). We have compared the sensitivity and specificity of ΔMD_{LAD} and $\Delta V_{25Gyheart}$ provided by the simple method based on a single CT scan with that of the original method that indicated high consistency [19] (Table 1). Based on these findings, the cut-off values of 0.6 Gy (ΔMD_{LAD}) and 1.0% ($\Delta V_{25Gyheart}$) have been selected for further analyses and practice.

Next, the concordance of calculator-predicted treatment position based on ΔMD_{LAD} *vs.* $\Delta V_{25Gyheart}$ and the need for intervention were analyzed in the validation set. In 28 supine-predicted cases and 64 prone-predicted cases, the same treatment position was suggested by both measures (Table 3). Among the 28 supine-predicted cases in 2, the radiotherapy plan revealed that MD_{LAD} > 12.5 Gy, but only 1 could be improved by changing the treatment position. Among the 64 prone-predicted cases in 8, the MD_{LAD} exceeded the dose constraint of 12.9 Gy; only 3 plans could be improved by applying the supine position. Among the discordant cases, ΔMD_{LAD} suggested prone position in 3 and supine position in 5 cases; in both groups in a single case each could the LAD dose be improved by changing the treatment position. In altogether 7 cases, a different intervention (IMRT) had to be applied (Table 3).

"Routine clinical practice" set

In the "routine clinical practice" series of 60 patients, the new method proved feasible. All patients received treatment in the position suggested by the calculator except one, who had to receive a second CT in the other position due to unacceptable LAD dose. The other patients had MD_{LAD} and $V_{25Gyheart}$ values well below the predefined dose limits, and these were similar to the values calculated in the validation set (Table 4).

External testing

In a series of 28 breast cancer patients from Liege, the predictors BMI, D_{med} and A_{heart} significantly differed from the same parameters among the patients from Szeged. In 18/28 cases, P_{ref} was equal or close to P_{med} ($\leq$6 mm), while in 10, cases P_{ref} varied from P_{med} by 9–16 mm. Comparing the calculator-provided dose differences with the treatment planning data, favored treatment position was correct in 24/28 (accuracy: 85.7%) and 23/28 (accuracy: 82.1%) cases taking into account the LAD and heart doses,

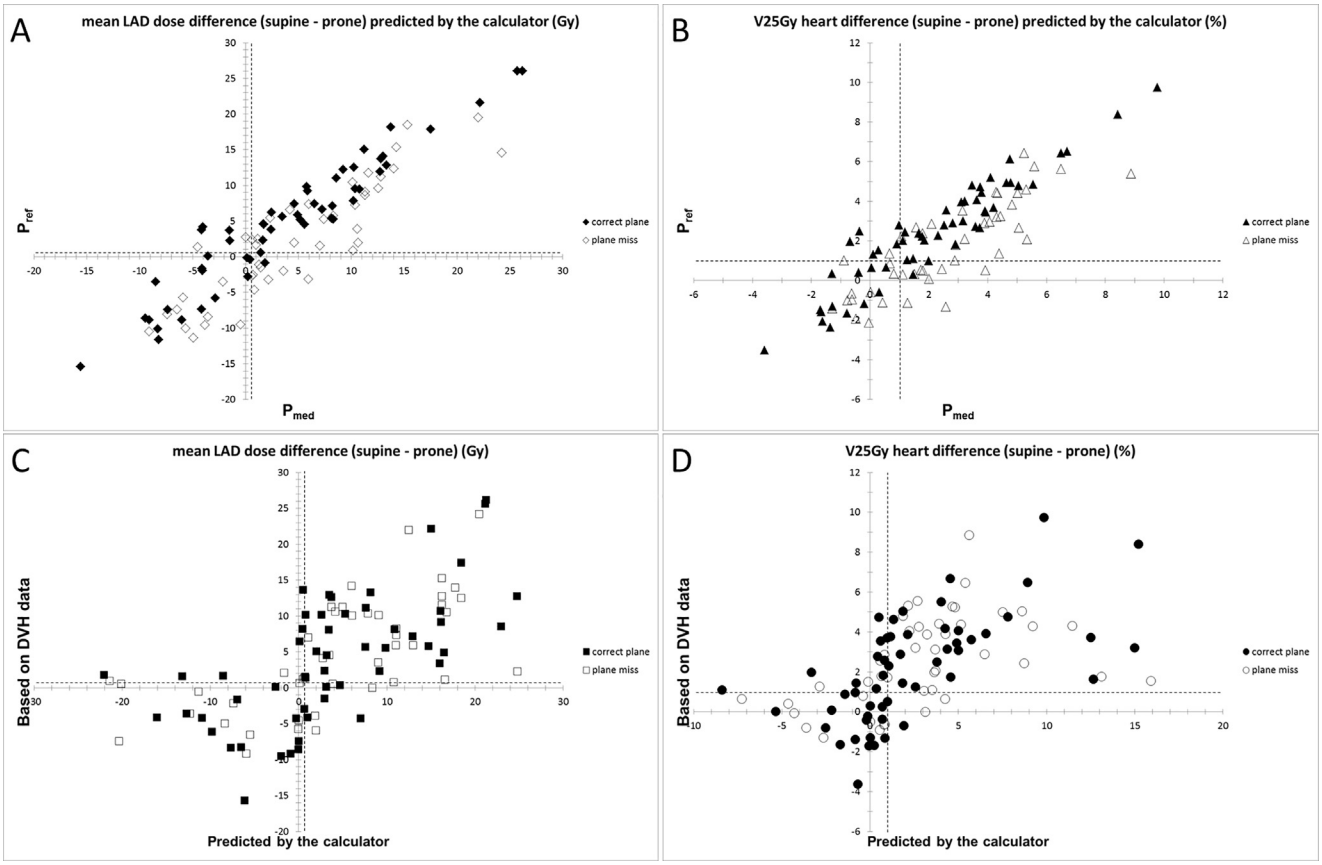


Fig. 2. Calculator suggestion of LAD (A) and heart (B) dose differences by the input of D_{med} and A_{heart} based on P_{ref} *vs.* P_{med}; LAD (C) and heart (D) doses according to the estimation of the simple clinical method based on a single CT scan *vs.* DVH data extracted from the planning system (n = 100). Dashed lines indicate the cut-off values of 0.6 Gy (ΔMD_{LAD}) and 1.0% ($\Delta V_{25Gyheart}$) specified by sensitivity and specificity values.

Table 3
Concordance of treatment position as predicted by ΔMD_{LAD} *vs.* $\Delta V_{25Gyheart}$, in the validation set (n = 100). In concordant cases the suggested position, in discordant cases the position suggested by ΔMD_{LAD} was applied unless the dose constraints were exceeded; in such cases the other treatment position or alternative techniques may be tested.

		$\Delta V_{25Gyheart}$							
		Supine				Prone			
		All	MD_{LAD} > 12.5 Gy	Change position	Other intervention	**All**	MD_{LAD} > 12.9 Gy	Change position	Other intervention
ΔMD_{LAD}	*Supine*	28	2	1/2	1/2	5	1/5	1/1	–
	Prone	3	2/3	1/2	1/2	64	8/64	3/8	5/8

Table 4
LAD and heart doses in the validation set and the "routine clinical practice" series: in the majority of cases, LAD and heart doses were well below the position-related dose constraints; for those patients who had higher than accepted doses, an alternative technique had to be applied.

	Treatment position	n (%)	Mean LAD dose (Gy)				$V_{25Gyheart}$ (%)			
			Mean	SD	Min	Max	Mean	SD	Min	Max
Validation series	*Prone*	67 (67.0)	6.55	6.03	1.70	26.66	1.16	2.24	0.0	8.75
	Supine	33 (33.0)	6.90	3.86	1.71	13.73	1.54	1.38	0.0	4.77
"Routine clinical practice" series	*Prone*	47 (78.3)	6.58	2.29	1.95	11.24	0.86	0.57	0.1	2.67
	Supine	13 (21.7)	7.35	3.05	2.54	15.85	1.15	0.95	0.21	3.57

respectively. Sensitivity and specificity of ΔMD_{LAD} was 83.3% and 86.4%, respectively, whereas sensitivity and specificity of $\Delta V_{25Gyheart}$ was 100.0% and 80.0%, respectively.

Discussion

According to the present study and others [11,14,15,19,20], in about 20% of the cases, prone positioning during left breast radiotherapy increases the dose to the LAD or the heart. To estimate and select the preferable positioning mode, supine CT seems the best approach to consider the patient's anatomical determinants. We have shown that a single CT scan at the middle of the heart may replace a whole CT series by providing consistent anatomical data thus avoiding extra radiation exposure to the patient and work load to the staff. Based on the outcome of the external implementation of the method on an independent case series, we recommend its use after local testing.

Our validated statistical model for predicting the preferable treatment position utilizes 3 specific measures, and seems the most complex predictive tool for this purpose in the literature [19]. In other studies, the in-field heart volume [16,17,25] and most frequently the size of the breast [4,11] have been used for selection. An increased BMI has also been related to larger heart doses [28] or consequential radiation cardiac morbidity [29], but its role in predicting benefit of prone positioning may be refined by the use of other patient-related parameters [19]. We consider the BMI in our calculator as a stable parameter while there is potential uncertainty in the specification of P_{ref} or imprecision in the actual measurement of D_{median} or A_{heart} on a given image. Nevertheless, detailed analysis indicates that accidental imprecision does not significantly influence final prediction (data not shown). The dose constraints optimized by individual positioning provides additional safety in practice. Despite the lack of full equivalence of the data extracted from the original method *vs.* the new method, the ultimate consistency still seems to qualify the developed "simple tool" for clinical application.

External use indicated similar accuracy as the originally developed method. Despite the reassuring results on an independent series of patients in a radiotherapy center using a slightly different protocol, the utility of the reported clinical tool could be compromised by the diversity of practice in others. PTV contouring depends on repositioning accuracy and the method of treatment verification. Interfractional differences may be especially large in the prone position [18,30]. Lakosi et al. found population systematic error values of 4.5/3.9/3.3 mm in the lateral/longitudinal/vertical directions, while the random error was 5.4/3.8/2.8 mm [27]. Among our recent breast radiotherapy cases, the population systematic and random error in the lateral/longitudinal/vertical directions was similar in the prone position *vs.* supine position (3.4/2.3/2.7 mm and 7.8/4.6/6.9 mm, respectively *vs.* 2.2/3.0/1.6 mm and 6.7/5.5/4.5 mm, respectively). Only some groups study the dose to the coronary arteries [11,12,19,20,31–34]. The outlining of the coronary vessels shows significant inter-observer variation that may jeopardize dose verification in the selected position [35,36]. Different approaches have been tested to improve consistency including the administration of contrast media [35–37]. Lee et al. developed a new protocol to outline the LAD region which included 96% of the LAD volume as delineated by 4 experienced radiation oncologists [37]. Significant impact was made by the implementation of specific guidelines [35–37]. Since the utility of the simple tool might be influenced by several factors, in addition to the use of institutional LAD contouring guidelines and study of inter-observer variation, we consider essential its testing before routine use. In the case of hypofractionated radiotherapy, the model parameters of the calculator should be re-estimated and the dose constraints should be re-defined.

The benefit of positioning prone *vs.* supine may be discordant by means of LAD and heart doses [11,19,34]. We regard the LAD dose as a surrogate indicator of radiation harm due to its proven role in late cardiac morbidity [3] and because the LAD being situated on the anterior surface of the heart is a sensitive marker of danger if the heart is at all included into radiation. Our strategy for optimization in individual cases is to consider the MD_{LAD} as priority that is usually confirmed by the heart dose (as was true for 92% of cases in our series).

The radiation exposure of the heart may be significantly reduced by the use of respiration-guided techniques including the deep inspiration breath hold (DIBH) technique and respiratory gating. In the UK HeartSpare study, supine DIBH provided superior cardiac sparing than a free-breathing prone position in large-breasted women [12]. Interestingly, the implementation of DIBH in the prone position gave the optimal heart sparing results as compared with that in the supine position or free-breathing [33]. There are some centers that due to resource limitations prioritize high cardiac dose cases for DIBH [38]. Our tool could be used for patients either not amenable for or not having access to DIBH due to patient-specific features (cardiorespiratory problems, lack of compliance) or limited/no resources, respectively.

We think that since a linear, no-threshold association exists between the mean heart dose and coronary events [3], doses to the LAD, right coronary artery or the circumflex artery should be controlled [20]. Nevertheless, the utilization of heart dose–volume data only is a possibility if LAD contouring cannot be afforded. Since good agreement exists between the mean heart dose and $V_{25Gyheart}$ (Rprone: 0.98, Rsupine: 0.99) or MD_{LAD} (Rprone and Rsupine: 0.87) in both positions ($p < 0.001$ in all comparisons), here the presented tool could be adapted to practices which adhere to the consideration of the mean heart dose.

In summary, we have demonstrated great consistency of our method based on a validated model for the prediction of treatment position prone *vs.* supine with less heart exposure during left breast radiotherapy; the simplified tool presented here omits the performance of planning CT in both positions. Based on the results of its external testing, we truly recommend its use in centers that apply prone positioning in routine clinical practice. Due to differences in populations and radiotherapy protocols, local testing is essential.

Conclusion

We consider this simple clinical tool useful for assisting individual positioning in routine clinical practice aiming at maximum heart protection during left breast irradiation.

Conflict of interest

None declared.

Acknowledgement

This work had been supported by the VKSZ 12-1-2013-0012 project. The authors are grateful to Edit Kiss and Laura Gál for their excellent technical support.

References

[1] Clarke M, Collins R, Darby S, et al. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). 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 2005;366:2087–106.
[2] McGale P, Darby SC, Hall P, et al. Incidence of heart disease in 35,000 women treated with radiotherapy for breast cancer in Denmark and Sweden. Radiother Oncol 2011;100:167–75.
[3] Darby SC, Ewertz M, McGale P, et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. N Engl J Med 2013;368:987–98.
[4] Taylor CW, Kirby AM. Cardiac side-effects from breast cancer radiotherapy. Clin Oncol (R Coll Radiol) 2015;7:621–9.
[5] Shah C, Badiyan S, Berry S, et al. Cardiac dose sparing and avoidance techniques in breast cancer radiotherapy. Radiother Oncol 2014;112:9–16.
[6] Correa CR, Litt HI, Hwang WT, Ferrari VA, Solin LJ, Harris EE. Coronary artery findings after left-sided compared with right-sided radiation treatment for early-stage breast cancer. J Clin Oncol 2007;25:3031–7.
[7] Harris EE, Correa C, Hwang WT, et al. Late cardiac mortality and morbidity in early-stage breast cancer patients after breast-conservation treatment. J Clin Oncol 2006;24:4100–6.
[8] Nilsson G, Holmberg L, Garmo H, et al. Distribution of coronary artery stenosis after radiation for breast cancer. J Clin Oncol 2012;30:380–6.
[9] Nilsson G, Witt Nyström P, Isacsson U, et al. Radiation dose distribution in coronary arteries in breast cancer radiotherapy. Acta Oncol 2016;55:959–63.
[10] Becker-Schiebe M, Stockhammer M, Hoffmann W, Wetzel F, Franz H. Does mean heart dose sufficiently reflect coronary artery exposure in left-sided breast cancer radiotherapy?: influence of respiratory gating. Strahlenther Onkol 2016;192:624–31.

[11] Kirby AM, Evans PM, Donovan EM, Convery HM, Haviland JS, Yarnold JR. Prone versus supine positioning for whole and partial-breast radiotherapy: a comparison of non-target tissue dosimetry. Radiother Oncol 2010;96:178–84.
[12] Bartlett FR, Colgan RM, Donovan EM, et al. The UK HeartSpare Study (Stage IB): randomised comparison of a voluntary breath-hold technique and prone radiotherapy after breast conserving surgery. Radiother Oncol 2015;114:66–72.
[13] Griem KL, Fetherston P, Kuznetsova M, et al. Three-dimensional photon dosimetry: A comparison of treatment of the intact breast in the supine and prone position. Int J Radiat Oncol Biol Phys 2003;57:891–9.
[14] Lymberis SC, DeWyngaert JK, Parhar P, et al. Prospective assessment of optimal individual position (prone versus supine) for breast radiotherapy: volumetric and dosimetric correlations in 100 patients. Int J Radiat Oncol Biol Phys 2012;84:902–9.
[15] Formenti SC, DeWyngaert JK, Jozsef G, Goldberg JD. Prone vs supine positioning for breast cancer radiotherapy. JAMA 2012;308:861–3.
[16] DeWyngaert JK, Jozsef G, Mitchell J, Rosenstein B, Formenti SC. Accelerated intensity-modulated radiotherapy to breast in prone position: dosimetric results. Int J Radiat Oncol Biol Phys 2007;68:1251–9.
[17] Formenti SC, Gidea-Addeo D, Goldberg JD, et al. Phase I-II trial of prone accelerated intensity modulated radiation therapy to the breast to optimally spare normal tissue. J Clin Oncol 2007;25:2236–42.
[18] Varga Z, Hideghéty K, Mező T, Nikolényi A, Thurzó L, Kahán Z. Individual positioning: a comparative study of adjuvant breast radiotherapy in the prone versus supine position. Int J Radiat Oncol Biol Phys 2009;75:94–100.
[19] Varga Z, Cserháti A, Rárosi F, et al. Individualized positioning for maximum heart protection during breast irradiation. Acta Oncol. 2014;53:58–64.
[20] Würschmidt F, Stoltenberg S, Kretschmer M, Petersen C. Incidental dose to coronary arteries is higher in prone than in supine whole breast irradiation. A dosimetric comparison in adjuvant radiotherapy of early stage breast cancer. Strahlenther Onkol 2014;190:563–8.
[21] Dundas KL, Pogson EM, Batumalai V, et al. Australian survey on current practices for breast radiotherapy. J Med Imaging Radiat Oncol 2015;59:736–42.
[22] Lakosi F, Gulyban A, Janvary L, et al. Respiratory motion, anterior heart displacement and heart dosimetry: comparison between prone (Pr) and supine (Su) whole breast irradiation. Pathol Oncol Res 2015;21:1051–8.
[23] Merchant TE, McCormick B. Prone position breast irradiation. Int J Radiat Oncol Biol Phys 1994;30:197–203.
[24] Buijsen J, Jager JJ, Bovendeerd J, et al. Prone breast irradiation for pendulous breasts. Radiother Oncol 2007;82:337–40.
[25] Zhao X, Wong EK, Wang Y, et al. A support vector machine (SVM) for predicting preferred treatment position in radiotherapy of patients with breast cancer. Med Phys 2010;37:5341–50.
[26] Feng M, Moran JM, Koelling T, et al. Development and validation of a heart atlas to study cardiac exposure to radiation following treatment for breast cancer. Int J Radiat Oncol Biol Phys 2011;79:10–8.
[27] Lakosi F, Gulyban A, Ben-Mustapha Simoni S, et al. Feasibility evaluation of prone breast irradiation with the Sagittilt© system including residual-intrafractional error assessment. Cancer Radiother 2016;20:776–82.
[28] Evans SB, Sioshansi S, Moran MS, Hiatt J, Price LL, Wazer DE. Prevalence of poor cardiac anatomy in carcinoma of the breast treated with whole-breast radiotherapy: reconciling modern cardiac dosimetry with cardiac mortality data. Am J Clin Oncol 2012;35:587–92.
[29] Evans ES, Prosnitz RG, Yu X, et al. Impact of patient-specific factors, irradiated left ventricular volume, and treatment set-up errors on the development of myocardial perfusion defects after radiation therapy for left-sided breast cancer. Int J Radiat Oncol Biol Phys 2006;66:1125–34.
[30] Mulliez T, Gulyban A, Vercauteren T, et al. Setup accuracy for prone and supine whole breast irradiation. Strahlenther Onkol 2016;192:254–9.
[31] Aznar MC, Korreman SS, Pedersen AN, Persson GF, Josipovic M, Specht L. Evaluation of dose to cardiac structures during breast irradiation. Br J Radiol 2011;84:743–6.
[32] Mulliez T, Veldeman L, van Greveling A, et al. Hypofractionated whole breast irradiation for patients with large breasts: a randomized trial comparing prone and supine positions. Radiother Oncol 2013;108:203–8.
[33] Mulliez T, Veldeman L, Speleers B, et al. Heart dose reduction by prone deep inspiration breath hold in left-sided breast irradiation. Radiother Oncol 2015;114:79–84.
[34] Verhoeven K, Swelden C, Petillion S, et al. Breathing adapted radiation therapy in comparison with prone position to reduce the doses to the heart, left anterior descending artery and contralateral breast in whole breast radiation therapy. Pract Radiat Oncol 2014;4:123–9.
[35] Lorenzen EL, Taylor CV, Maraldo M, et al. Inter-observer variation in delineation of the heart and left anterior descending coronary artery in radiotherapy for breast cancer: a multi-centre study from Denmark and the UK. Radiother Oncol 2013;108:254–8.
[36] Duane F, Aznar MC, Bartlett F, et al. A cardiac contouring atlas for radiotherapy. Radiother Oncol 2017;122:416–22.
[37] Lee J, Hua K-L, Hsu S-M, et al. Development of delineation for the left anterior descending coronary artery region in left breast cancer radiotherapy: an optimized organ at risk. Radiother Oncol 2017;122:423–30.
[38] Tanna N, McLauchlan R, Karis S, Welgemoed C, Gujral DM, Cleator SJ. Assessment of upfront selection criteria to prioritise patients for breath-hold left-sided breast radiotherapy. Clin Oncol (R Coll Radiol) 2017;29:356–61.