

RESEARCH

Open Access



A comparative study of different parameter estimation methods for predictive models of Normal Tissue Complication Probability (NTCP) of radiation-induced temporal lobe injury following intensity-modulated radiotherapy in nasopharyngeal carcinoma

Huidan OuYang^{1†}, Yuze Liu^{2†}, Xianming He³, Jianze Zhang³, Lei Tao³, Mengmeng Liu³, Jianwu Ding³, Ronghuan Hu³, Jiali Hu³, Zequn Huang³, Su Deng⁵, Jiayin Wu⁵, Zhengyu Xu^{4*}, Qiwei Luo^{3*} and Lei Zeng^{3*}

Abstract

Background Normal Tissue Complication Probability (NTCP) models predict temporal lobe injury risk post-intensity-modulated radiotherapy in nasopharyngeal carcinoma patients. Optimal parameter estimation methods for NTCP models need refinement.

Purpose To identify optimal method for parameter estimation in Normal Tissue Complication Probability models for temporal lobe injury following intensity-modulated radiotherapy in nasopharyngeal carcinoma patients.

Materials and methods In this study, all patients underwent curative intensity-modulated radiation therapy at two research centers. Data of temporal lobes from three cohorts [Data-A, $n = 278$ (training set); Data-B, $n = 119$ (external validation set); Data-C, $n = 215$ (internal validation set)]. Five NTCP models were considered, including the Serial Reconstruction Unit (SRU) model, Poisson model, Lyman model, Logit model and Logistic model. Three parameter estimation methods, namely Bayesian estimation (BE), Least Squares Estimation (LSE) and Maximum Likelihood Estimation (MLE), were applied to calibrate the five NTCP models. Area Under Curve (AUC), confusion matrices, dose-response curves were used to compare the performance of the models.

Results Six hundred twelve patients were enrolled, with 278 in the Data-A; 119 in the Data-B; 215 in the Data-C. The Poisson-NTCP model was evaluated using AUC and R^2 values across three parameter estimation methods (BE, LSE, and MLE) on three datasets. The results were as follows: Data-A: BE (AUC: 0.938, R^2 : 0.953), LSE (0.942, 0.986), MLE

[†]Huidan OuYang and Yuze Liu made equal contributions to this study.

*Correspondence:

Zhengyu Xu

jyh@qq.com

Qiwei Luo

luoqiwei@email.ncu.edu.cn

Lei Zeng

ndefy19019@ncu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

(0.940, 0.843); Data-B: BE (0.744, 0.958), LSE (0.743, 0.697), MLE (0.745, 0.857); Data-C: BE (0.867, 0.915), LSE (0.862, 0.916), MLE (0.865, 0.896). Compared with the remaining models, the Poisson-NTCP model based on BE had also better performance of fitting dose–response curve and recall rate, accuracy and specificity of confusion matrix.

Conclusion Bayesian Estimation (BE) is the best parameter estimation method among the three parameter estimation methods. The Poisson-NTCP model based on BE exhibited the best fit to the data in predicting post-IMRT incidence of TLI in NPC.

Keywords Parameter Estimation, Normal Tissue Complication Probability, Nasopharyngeal Carcinoma

Introduction

Radiotherapy, particularly intensity-modulated radiotherapy, is a primary treatment for malignant tumors with dosimetric advantages over traditional techniques. In previous studies, numerous scholars have proposed various Normal Tissue Complication Probability (NTCP) models. For instance, Momeni et al. used the Lyman-Kutcher-Burman (LKB) model to predict the NTCP of acute eyelid erythema in patients with head and neck cancers and skull-base tumors after radiotherapy [1]. Lai et al. used a modified generalized Lyman normal-tissue complication probability model to estimate the risk of major adverse cardiac events after radiotherapy for left-sided breast cancer [2]. Rancati et al. utilized the logit-EUD model combined with clinical risk factors to predict the occurrence of late toxicities in rectal cancer patients after radiotherapy [3]. Wang et al. used the LKB model to predict the occurrence of $\geq$ grade 2 hematological toxicity after radiotherapy for cervical cancer [4]. Jackson applied the parallel model to analyze the NTCP of radiation-induced hepatitis [5]. Van Dijk LV et al. developed an NTCP model for osteoradionecrosis of the mandible in patients with head and neck cancer after radiation therapy [6], and Gloi used the Poisson-EUD model to evaluate the NTCP of a new partial breast irradiation method, MammoSite RTS [7].

Predecessors used the maximum likelihood method for estimating parameters in their models. However, this method is unsuitable when dealing with non-normally distributed data, or correlated data, leading to estimation issues or calculation failures [8, 9]. In addition, the parameter estimation method for a single model limits the model fitting to a local optimum, resulting in larger errors in the outcomes [10].

In previous NTCP-related studies, five models were used: Lyman, Logit, SRU, Poisson, and Logistic, with 3, 3, 2, 2, and 3 parameters, respectively. Our earlier research on temporal lobe injury complications post-IMRT in nasopharyngeal carcinoma patients showed the NTCP curve is S-shaped, ranging from 0 to 1. For these models, we plan to use three parameter estimation methods: maximum likelihood, least squares, and Bayesian estimation. Summarizing the research achievements of

predecessors, they all directly used the maximum likelihood method to estimate the parameters. For example, Tucker et al. compared the fitting of various NTCP models with the late rectal toxicity data of 128 prostate cancer patients and determined that the mean dose model is the best model for predicting late rectal injury after irradiation [11]. Semenenko et al. obtained parameter estimates for the LKB model through a comprehensive analysis of lung (radiation pneumonitis) and parotid gland (xerostomia) toxicity data from multiple institutions [12]. Momeni et al. used the generalized LKB model to determine the NTCP of acute ocular pain after radiotherapy for head and neck cancer, and the study suggested that keeping the average dose of the eyeball below 25 Gy, the probability of ocular pain will be less than 12% [13]. D'Avino et al. compared the predictive capabilities of the multivariate Logistic model and the LKB model for gastrointestinal toxicity after radiotherapy for localized prostate cancer, indicating that the predictive performance of the multivariate model is better than that of the LKB model [14]. Chapet used the Lyman model to predict acute esophagitis NTCP post-radiotherapy for non-small cell lung cancer [15]. However, few studies have compared the goodness of fit of different parameter estimation methods for models. This study, using data from two centers and three groups, compared parameter estimation methods' impact on accuracy via AUC and R^2 , aiming to optimize NTCP model parameters for improved radiotherapy planning.

Materials and methods

The study was approved by the Biomedical Research Ethics Committee of the Second Affiliated Hospital of Nanchang University (No. O-2023107). In this study, all computer code used for modeling and/or data analysis has been uploaded to the GitHub repository, which is available at <https://github.com/yxouyang/NTCP-software>. All data were analyzed using Python (version 3.7).

Data collection

The study population consisted of 612 previously untreated nasopharyngeal carcinoma patients without

metastasis from two academic institutions. A total of 278 patients were recruited from the Cancer Center of Sun Yat-sen University (Data-A) from January 2003 to February 2008 formed the training dataset for model development, and 215 patients were recruited from the same center (Data-C) from January 2012 to June 2012 constituted the internal validation dataset. Additionally, 119 patients were from the Second Affiliated Hospital of Nanchang University from January 2016 to May 2018 formed the external validation dataset (Data-B). Inclusion criteria for patients in this study were as follows: pathologically confirmed NPC without distant metastasis; initially treated with definitive IMRT; complete DVH data available; and a follow-up period exceeding 60 months on magnetic resonance imaging (MRI) or MRI-based diagnosis of TLL. Clinical characteristics and dosimetric features of the patients were collected; clinical features included age, gender, TNM staging, chemotherapy and radiotherapy regimens, radiation doses and fractions, the site of radiation-induced temporal lobe injury, diagnosis time, survival time, etc.; dosimetric features included Dose Volume Histogram (DVH) parameters for each temporal lobe, maximum dose (Dmax), absolute volume, relative volume, dose delivered to a 0.5-cm³ volume (D0.5 cc), and dose delivered to a 1-cm³ volume (D1cc), etc.

Conversion of physical data to biological data

The collected data were physical in nature, and for different radiotherapy fractions, it was necessary to convert them into biological data to standardize the measurements. Dose volume histograms were rescaled for a treatment schedule of 2 Gy per fraction by using a linear quadratic model and an α/β value of 3 Gy for biologic end point late toxicity in normal temporal lobe. The formula is as follows:

$$D_2 = D_x(\alpha/\beta + dx)/(\alpha/\beta + 2) \quad (1)$$

where D_x is the total dose delivered to the x -th fraction, and D_2 is the equivalent uniform dose per 2 Gy.

Calculation of EUD

Different dose-volume histogram (DVH) reduction schemes have been used to define the summary measure μ , such as the effective volume or effective dose in the LKB model. Here, the generalized equivalent uniform dose (EUD) is defined as the dose's Lebesgue norm according to the following power law relationship

$$\mu = \text{EUD} = (\sum_i v_i D_i^a)^{1/a} \quad (2)$$

The sum is calculated over all bins (v_i , D_i) of the differential DVH, and a is a parameter describing the

dose-volume effect. When $a = \infty$ (i.e., no volume effect), EUD equals the maximum dose. For a with a value of 1, Eq. (2) gives the mean dose (large volume effect).

For non-uniform dose distribution, it can be defined as the equivalent uniform dose that, when applied uniformly to the entire organ, would give the same macroscopic dose-response ($V=1$). It can then be defined as:

$$\text{EUD}_{\text{SRU}} = \frac{1}{\sigma} \log(\sum_i v_i \exp(\delta D_i)) \quad (3)$$

NTCP models

Lyman-EUD model

In the Lyman-EUD model, the parameter s describes the slope of the sigmoidal response curve at the steepest point $\mu = \mu_{50}$. The NTCP function predicts a 50% complication probability. Typically, the slope parameter s is replaced by its inverse m , according to $s = 1/(m \cdot \mu_{50})$.

$$\text{NTCP}_{\text{Lyman}}(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{s(\mu - \mu_{50})} \exp(-x^2/2) dx \quad (4)$$

Lyman-EUD model has three parameters a , m , and μ_{50} that need to be fitted.

Logit-EUD model

The logit-EUD model also uses the generalized EUD Eq. (2) as a summary measure μ . Its two parameters μ_{50} and k are determined by EUD. Therefore, with the addition of the EUD parameter a , the model has three parameters

$$\text{NTCP}_{\text{Logit}}(\mu) = \frac{1}{1 + (\mu_{50}/\mu)^k} \quad (5)$$

Serial Reconstruction Unit (SRU) model

The Serial Reconstruction Unit model, recently proposed by Alber and Belka, describes radiation complications as the failure of a dynamic repair process. In this model, σ is the organ-specific sensitivity parameter, and D_0 is the reference dose.

$$\text{NTCP}_{\text{SRU}}(V, D) = 1 - \exp(-V \exp(\sigma(D - D_0))) \quad (6)$$

For inhomogeneous dose distributions an equivalent uniform dose, which would give the same macroscopic dose-response when applied homogeneously to the whole organ ($V=1$), can be defined as Eq. (3). Therefore, the NTCP function is given by:

$$\text{NTCP}_{\text{SRU}}(\text{EUD}_{\text{SRU}}) = 1 - \exp(-\exp(\sigma(\text{EUD}_{\text{SRU}} - D_0))) \quad (7)$$

In this study, the Serial Reconstruction Unit (SRU) model is described by two parameters: σ and $D_{0\sigma}$.

Poisson-EUD model

Similar to the SRU model, the Poisson-EUD model uses a mechanism concept to describe the primary serial tissue dose response. Assuming complications are the result of local dose responses of non-interacting subunits, the following NTCP function can be derived based on Poisson statistics.

$$\text{NTCP}_{\text{Poisson}}(\text{EUD}) = 1 - \exp\left[-\left(\frac{\text{EUD}}{D_0}\right)^a\right] = 1 - \exp\left[-\ln 2 \left(\frac{\text{EUD}}{D_{50}}\right)^a\right] \quad (8)$$

With a reference dose D_0 (or a dose D_{50} causing 50% complication probability) and a volume-effect (steepness) parameter a . The EUD is given by Eq. (2), where, according to this model, the exponent of the Equivalent Uniform Dose (EUD) and the steepness parameter of the NTCP function share the same value. Consequently, unlike the Lyman-EUD and Logit-EUD models, the Poisson-EUD model is characterized by only two parameters.

Logistic regression model

In the logistic regression model, β_0 (intercept) is a constant and β_1 is the logistic regression coefficients of the predictor variable, respectively. The logistic regression model also utilizes the generalized EUD Eq. (2). Again, μ serves as a comprehensive evaluation index. Consequently, in conjunction with the parameter 'a' from the EUD, this model possesses three parameters altogether.

$$\text{NTCP}_{\text{Logistic}}(\mu) = \frac{1}{1 + \exp(-\beta_0 - \beta_1\mu)} \quad (9)$$

We established NTCP models using the model parameters obtained from Data-A dataset in both Data-B dataset and Data-C dataset, respectively.

Parameter estimation methods

Maximum likelihood estimation

The NTCP model is fitted to the data using the method of maximum likelihood.

$$L = \prod_{i=1}^N L_i p_i^{ep_i} (1 - p_i)^{(1-ep_i)} \quad (10)$$

This method determines the values of the model parameters that maximize the likelihood (L). This implies that the maximum value of L defines the greatest consistency between the true observed endpoints (ep_i) and the calculated NTCP values (p_i). The parameter N represents the number of data points. Mathematically, taking the logarithm facilitates faster convergence of the function.

$$\text{LL} = \ln(L) = \sum_{i=1}^N (ep_i \ln(p_i) + (1 - ep_i) \ln(1 - p_i)) \quad (11)$$

By automatically adjusting the parameters to maximize the natural logarithm of the likelihood ($\ln(L)$), the model is fitted. To express the uncertainty of the fitted parameters, the confidence intervals of the estimates are calculated using the profile likelihood method.

Bayesian parameter estimation

1. Determine the prior distribution of $p\{\theta\}$ of the parameter θ
2. From the sample set $D = \{x_1, x_2, \dots, x_n\}$, derive the joint distribution $p(D|\theta)$, which is a function of θ :

$$p(D|\theta) = \prod_{n=1}^N p(x_n|\theta) \quad (12)$$

3. Using Bayes' theorem, calculate the posterior distribution of θ :

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta} \quad (13)$$

4. Derive the Bayesian estimate:

$$\hat{\theta} = \int_{\theta} \theta p(\theta|D) d\theta \quad (14)$$

Least squares parameter estimation

That is, the sum of the squares of the differences between all observed values (samples) Y_i of the explained variable and the estimated value $\hat{\beta}_0 + \hat{\beta}_1 X_i$ is minimized (least squares), with the formula as follows:

$$\text{MinQ} = \sum_{i=1}^n [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)]^2 \quad (15)$$

Bayesian estimation differs significantly from traditional Maximum Likelihood Estimation (MLE) and Least Squares (LS). Traditional MLE estimates model parameters by finding the parameter values that maximize the likelihood function. It assumes that the data are independently and identically distributed and is sensitive to extreme values. When dealing with data close to 0 or 1 in probability, it may encounter computational anomalies. Least Squares, on the other hand, determines parameters by minimizing the sum of squared differences between the observed values and the estimated values, assuming that the data errors follow a normal distribution. Bayesian estimation, however,

is based on Bayes' theorem and combines prior knowledge with sample data to infer the posterior distribution of parameters. In clinical modeling, the advantage of Bayesian estimation lies in its ability to incorporate prior knowledge from clinicians, such as expected parameter ranges based on previous research or clinical experience. This prior information helps the model more reasonably estimate parameters, especially when data is limited. Additionally, Bayesian estimation better handles uncertainty, providing a more comprehensive evaluation of parameter uncertainty through the posterior distribution, which offers richer information for clinical decision-making.

Evaluation metrics

The performance of the models was assessed using R-squared (R^2) and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve as indicators.

R^2 : Defined for a set of observed values y_i and predicted values $\hat{y}_i$, R^2 is calculated as follows:

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (16)$$

The confusion matrix is presented in Table 1.

The AUC: A performance metric for the classifier, the AUC is calculated by integrating the area under the ROC curve, which plots the true positive rate against the false positive rate at various threshold settings:

$$AUC = \frac{1}{2} \sum_{i=1}^{m-1} (x_{i+1} - x_i)(y_i + y_{i+1}) \quad (17)$$

Results

Clinical features and EUD boxplot analysis

The clinical characteristics of the training and validation sets are presented in Table 2. The study utilized three datasets, Data-A, Data-B, and Data-C, for modeling and validation. Figure 1 illustrates the distribution of EUD across the five NTCP models. The central line within the box represents the median, the upper and lower bounds of the box denote the upper and lower quartiles, respectively, and the whiskers extending from the box represent the maximum and minimum EUD values. The width of the box provides an indication of the variability in EUD. Notably, Data-A exhibited a greater number of outliers, suggesting a higher degree of variability; Data-B displayed a more concentrated distribution with less variability, indicating greater stability; Data-C exhibits large volatility and is the most unstable.

Table 1 Confusion matrix: True Positive(TP); False Positive(FP); False Negative(FN); False Negative(FN)

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

Table 2 Clinical features of the training cohort and validation cohorts

	Data-A(n = 278)	Data-B(n = 119)	Data-C(n = 215)	p
Sex				0.068
Male	210(75.5%)	84(70.6%)	154(71.6%)	
Female	68(24.5%)	35(29.4%)	61(28.4%)	
Age				< 0.001
< 50	207(74.5%)	51(42.9%)	156(72.6%)	
≥ 50	71(25.5%)	68(57.1%)	59(27.4%)	
CCRT				< 0.001
Yes	186(66.9%)	113(95.0%)	159(74.0%)	
No	92(33.1%)	6(5.0%)	56(26.0%)	
T stage				< 0.0001
1	66(23.7%)	3(2.5%)	15(7.0%)	
2	60(21.6%)	19(16.0%)	43(20.0%)	
3	96(34.5%)	80(67.2%)	105(48.8%)	
4	56(20.2%)	17(14.3%)	52(24.2%)	
N stage				< 0.001
N0	69(24.8%)	1(0.8%)	19(8.8%)	
N1	134(48.2%)	6(5.1%)	86(40.0%)	
N2	56(20.2%)	109(91.6%)	91(42.4%)	
N3	19(6.8%)	3(2.5%)	19(8.8%)	

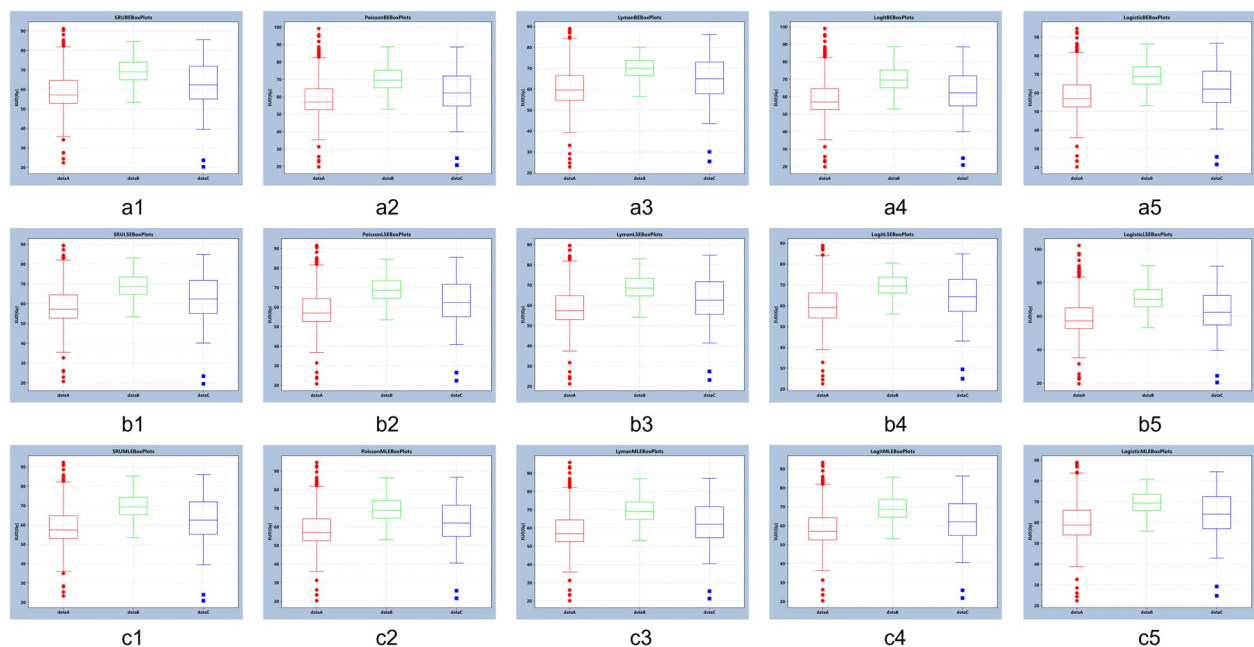


Fig. 1 Boxplots representing the parameter estimates obtained using Bayesian Estimation (BE), Least Squares Estimation (LSE), and Maximum Likelihood Estimation (MLE) (a-c) for the Serial Reconstruction Unit (SRU) model, Poisson model, Lyman model, Logit model, and Logistic model(1–5), respectively

Receiver Operating Characteristic (ROC) curve analysis

The area under the ROC curve (AUC) was employed as an evaluative metric to assess the accuracy of model predictions. An AUC value closer to 1 indicates a better model performance.

Figure 2, Fig. E1, and Fig. E2 present the ROC curves derived from the three datasets. The curves labeled B1-B5, L1-L5, and M1-M5 correspond to parameter estimations using Bayesian, Least Squares, and Maximum Likelihood methods, respectively. The results indicate that the AUC values are nearly identical across the three parameter estimation methods, suggesting that the models perform comparably regardless of the method used for parameter estimation.

Confusion matrix analysis

This study employed a confusion matrix to assess the performance of the NTCP model algorithms, with results based on the three different datasets shown in .

Figure 3, Fig.E3, and Fig.E4. The confusion matrices for Bayesian, Least Squares, and Maximum Likelihood estimations are denoted as B1-B5, L1-L5, and M1-M5, respectively. The color intensity of the diagonal blocks represents the difference between predicted and actual probabilities. The comparison of confusion matrix figures reveals that regardless of whether the data is from the experimental, internal validation, or external validation datasets, Bayesian estimation is more suitable for the

Poisson model, Least Squares estimation shows a good comparable performance between the Logit model and Lyman model, and Maximum Likelihood estimation is more effective for the Lyman model.

Incidence of temporal lobe injury and dose–response curves

Figure 4, Fig.E5, and Fig.E6 depict the incidence of temporal lobe injury and the corresponding dose–response curves based on the three datasets using the three different parameter estimation methods. The red and green crosses represent patients with and without temporal lobe injury, respectively. The vertical error bars indicate the 68% confidence intervals, and the blue data boxes shows the ratio of temporal lobe injured patients in the given dose for this patient group. As the Equivalent Uniform Dose (EUD) increases, the 68% confidence interval expands. The results demonstrate that the effects of the three parameter estimation methods are comparable and that there is an S-shaped curve relationship between NTCP and dose–response.

R² evaluation metrics

The R² values obtained after parameter estimation using BE, LSE, and MLE for the SRU, Poisson, Lyman, Logit, and Logistic models are presented in Table 3. The results indicate that the estimation effect of the Least Squares parameter estimation is similar to that

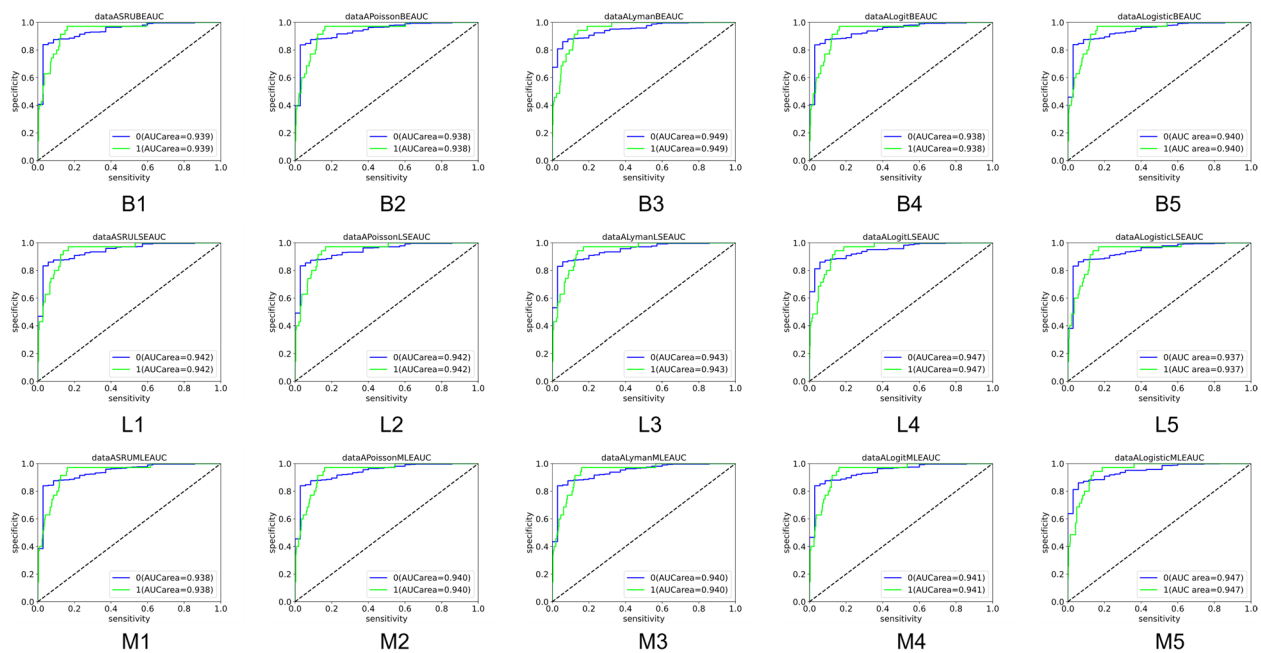


Fig. 2 The Receiver Operating Characteristic (ROC) curve based on the experimental dataset Data-A for the Serial Reconstruction Unit (SRU) model, Poisson model, Lyman model, Logit model, and Logistic model(1–5), respectively. B1–B5, L1–L5, and M1–M5 correspond to parameter estimations using Bayesian, Least Squares, and Maximum Likelihood methods

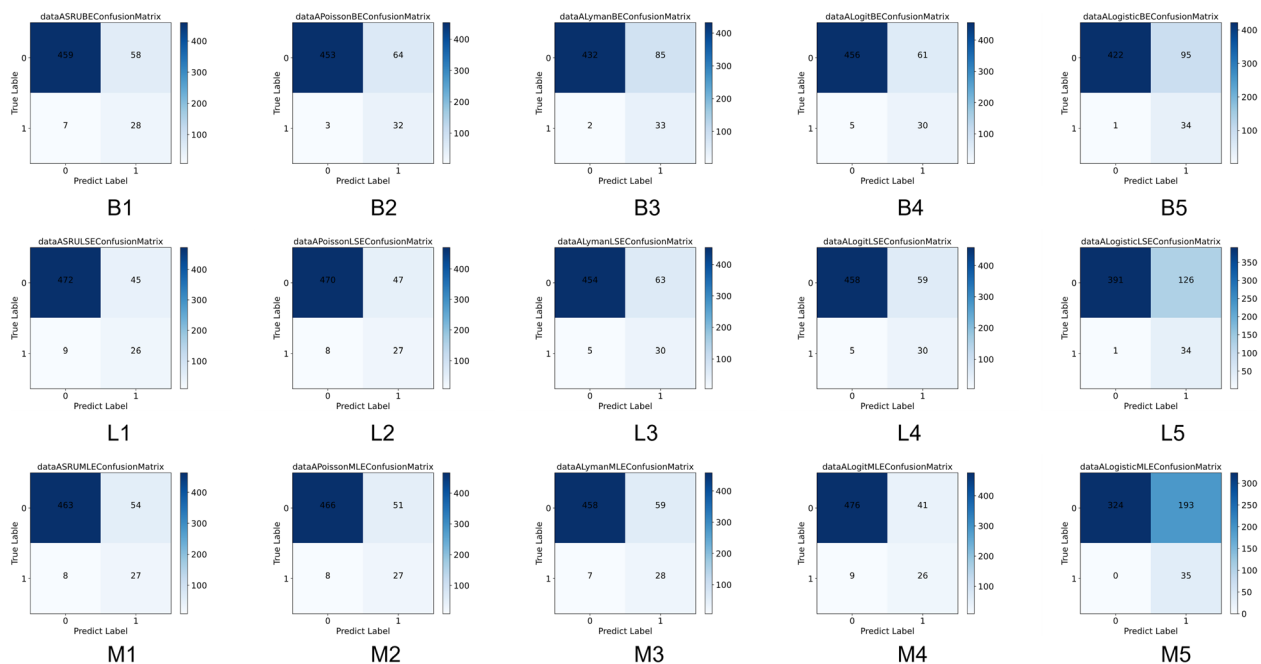


Fig. 3 Confusion matrix based on the experimental dataset Data-A for Bayesian, Least Squares, and Maximum Likelihood estimations are denoted as B1–B5, L1–L5, and M1–M5, respectively

of the Maximum Likelihood method, and in the construction of the Poisson model with Data-A, Data-B, and Data-C, the results obtained using Bayesian

estimation, Least Squares estimation, and Maximum Likelihood estimation are (BE: 0.953, 0.958, 0.915), (LSE: 0.986, 0.697, 0.916), and (ML: 0.843,

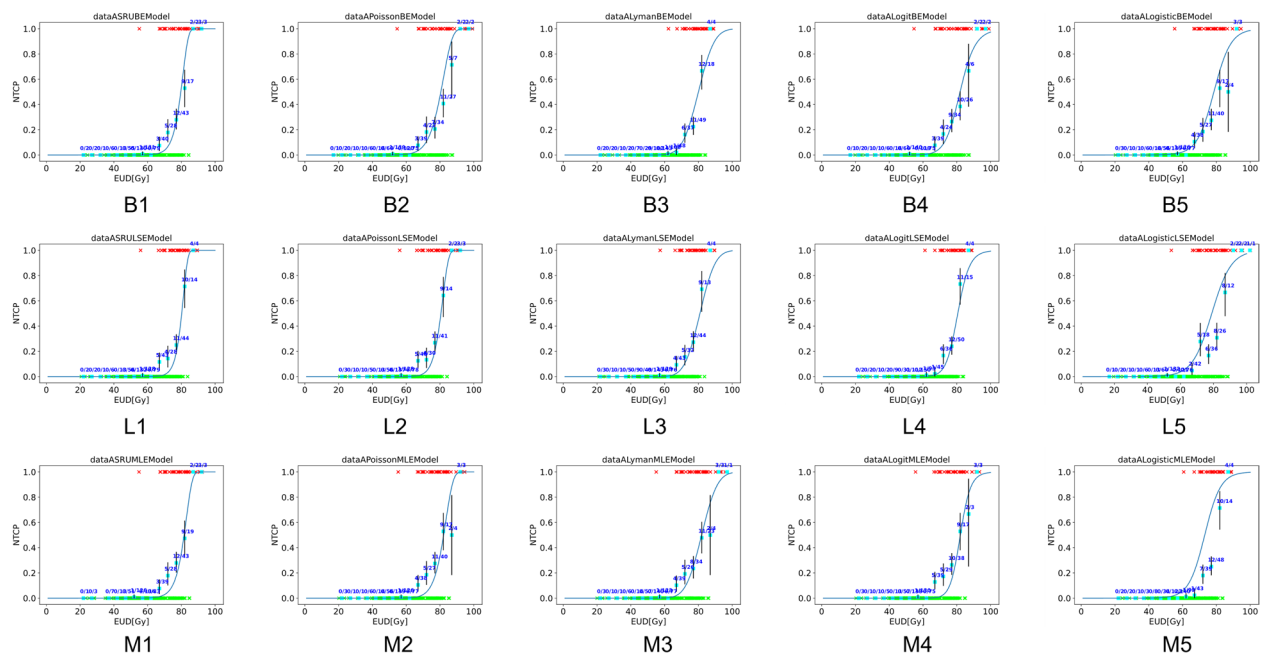


Fig. 4 The Normal Tissue Complication Probability (NTCP) and dose–response curves based on the experimental dataset Data-A for Bayesian, Least Squares, and Maximum Likelihood estimations are denoted as B1-B5, L1-L5, and M1-M5, respectively

Table 3 The regression models obtained after parameter estimation using Bayesian Estimation (BE), Least Squares Estimation (LSE), and Maximum Likelihood Estimation (MLE) for the Serial Reconstruction Unit (SRU) model, Poisson model, Lyman model, Logit model, and Logistic model. The R-squared (R2) statistic is utilized to assess the degree of fit of the regression models to the actual data

		SRU	Poisson	Lyman	Logit	Logistic
Data-A	BE	0.963	0.953	0.959	0.977	0.951
	LSE	0.977	0.986	0.907	0.963	0.920
	MLE	0.977	0.843	0.949	0.940	0.731
Data-B	BE	0.806	0.958	0.742	0.864	0.918
	LSE	0.762	0.697	0.679	0.671	0.956
	MLE	0.536	0.857	0.868	0.501	0.780
Data-C	BE	0.962	0.915	0.964	0.946	0.971
	LSE	0.837	0.916	0.876	0.890	0.923
	MLE	0.915	0.896	0.897	0.853	0.559

0.857, 0.896), respectively. For the Logistic model, the results are (BE: 0.951, 0.918, 0.971), (LSE: 0.920, 0.956, 0.923), and (ML: 0.731, 0.780, 0.559), respectively. For the Lyman model, the results are (BE: 0.959, 0.742, 0.964), (LSE: 0.907, 0.679, 0.876), and (ML: 0.949, 0.868, 0.897), respectively. It is evident that the Bayesian method shows superior performance in the Poisson model, the Least Squares estimation method has a comparable good performance with Bayesian method in the Logistic model, and the Maximum Likelihood method is possibly more suitable for the Lyman model.

Discussion

Parameter estimation methods in radiotherapy outcomes research, traditionally, the maximum likelihood method has been the cornerstone for estimating parameters in NTCP models post-radiotherapy. The prerequisite for using maximum likelihood estimation is that the data are approximately normally distributed and independent. However, for the everyone’s temporal lobes, there is some correlation between the left and right temporal lobes. If this data is used for NTCP modeling, it will inevitably lead to increased errors in the results. This study

introduces Bayesian estimation and least squares estimation as alternative methods for parameter estimation, providing a comparative analysis that reveals the least squares method's comparable efficacy to the maximum likelihood method. Bayesian estimation showed satisfactory results in this context.

In our previous research, the NTCP model constructed using the logistic algorithm achieved promising results [16]. The present study aimed to employ a variety of algorithmic models for comparative experiments to explore whether differences exist in the parameter estimation methods across different NTCP algorithmic models. The experimental results demonstrated that in the construction of the Poisson model with Data-A, Data-B, and Data-C, the R^2 values obtained using Bayesian estimation, least squares estimation, and maximum likelihood estimation were (BE: 0.953, 0.958, 0.915), (LSE: 0.986, 0.697, 0.916), and (ML: 0.843, 0.857, 0.896), respectively. It is evident that the Bayesian estimation method outperformed the least squares and maximum likelihood methods in the Poisson model. Conversely, in the construction of the Logistic model with Data-A, Data-B, and Data-C, the R^2 values obtained using Bayesian estimation, least squares estimation, and maximum likelihood estimation were (BE: 0.951, 0.918, 0.971), (LSE: 0.920, 0.956, 0.923), and (ML: 0.731, 0.780, 0.559), respectively. The Least Squares estimation method has comparable good performance with Bayesian method in the Logistic model. When constructing the Lyman model with Data-A, Data-B, and Data-C, the R^2 values from Bayesian estimation, least squares estimation, and maximum likelihood estimation were (BE: 0.959, 0.742, 0.964), (LSE: 0.907, 0.679, 0.876), and (ML: 0.949, 0.868, 0.897), respectively. The maximum likelihood estimation method demonstrated better performance in the Lyman model compared with the Bayesian and least squares methods. This indicates that the application of different parameter estimation methods to different NTCP algorithmic models is also a question worthy of investigation.

This study employed three datasets from two research centers for control experiments. This approach is instrumental in validating the generalizability of the parameter estimation methods. Among the three datasets, two were derived from the same research center but collected at different time points, allowing for internal validation. In contrast, the dataset from the other research center serves as an external validation. The collective validation using three datasets effectively mitigates the pitfalls of relying on a single dataset, which can lead to local optima and poor generalizability.

A rigorous comparison of parameter estimation methods necessitates the adoption of unbiased evaluation metrics that comprehensively quantify model accuracy,

generalizability, and robustness to ensure objective performance assessment. This study employs two evaluation metrics: the area under the receiver operating characteristic curve (AUC) and the coefficient of determination (R^2). AUC quantifies the model's ability to discriminate between classes (higher values indicating stronger performance), while R^2 measures the proportion of variance explained by the model in regression tasks (closer to 1 representing a strong goodness-of-fit). In the present study, the results indicate that the AUC values are nearly identical across the three parameter estimation methods, suggesting that the models perform comparably regardless of the method used for parameter estimation. However, it is evident that the Bayesian method shows superior performance in the Poisson model, the Least Squares estimation method has comparable good performance with Bayesian method in the Logistic model, and the Maximum Likelihood method is possibly more suitable for the Lyman model based on R^2 values. R -squared (R^2) is superior to AUC because R^2 measures the model's goodness of fit and the proportion of variance explained by the model in regression tasks. Historically, the maximum likelihood method has been used for estimating NTCP model parameters, with the Akaike Information Criterion (AIC) as the evaluation metric. However, AIC is clearly not suitable for least squares estimation and Bayesian parameter estimation, revealing its limitations. In addition to AUC and R^2 , there are other evaluation metrics such as F1 score, precision, accuracy, and Z-score, but ultimately, they share the same underlying logic, differing only in perspective. Utilizing two evaluation metrics to compare different parameter estimation methods helps to eliminate the randomness associated with single-indicator assessments.

Modern radiotherapy planning has evolved from physicians relying on guidelines and clinical experience to an approach that increasingly depends on algorithmic models based on volumetric dose data. In this research, numerous factors must be considered, such as the desirability for more and diverse dose-volume data, the applicability of NTCP algorithmic models, the precision of parameter estimation methods, and the computation of the Equivalent Uniform Dose (EUD) model. These factors require comprehensive consideration and continuous experimental validation to assist physicians in formulating more precise radiotherapy plans. This study focuses solely on comparing different parameter estimation methods, including maximum likelihood estimation, Bayesian parameter estimation, and least squares estimation. In reality, there are many estimation methods available, and with the advancement of precision medicine, these studies will become increasingly in-depth and meticulous.

To validate the effectiveness of the biophysical model, we conducted validation experiments predicting normal tissue complication rates under different radiation doses and fractionation schemes, and comparing the results with actual clinical data. The results showed a good consistency indicator (R^2 values) between the model-predicted complication rates and the actual occurrence rates, indicating that the biophysical model can accurately reflect the relationship between radiation dose and normal tissue complications to some extent.

From a clinical translation perspective, the NTCP model and parameter estimation method in this study have certain application potential. By accurately estimating the probability of normal tissue complications, clinicians can more reasonably adjust radiation doses and fractionation schemes during the radiation treatment planning phase, thereby reducing the risk of normal tissue complications. However, before clinical application, further validation in larger-scale multi-center clinical trials is necessary. The model should also be optimized and adjusted in consideration of clinical realities, such as patient individual differences, variations in radiation equipment to ensure its clinical practicality and reliability.

Conclusion

Our study affirms the efficacy of the Bayesian parameter estimation method in modeling the probability of normal tissue complications following tumor radiotherapy. Through comparative experiments with maximum likelihood estimation and least squares parameter estimation, it further demonstrates its superiority in the Poisson model. By utilizing data from research centers of different institutions, the generalizability of the parameter estimation method is evidenced, which will enhance the effectiveness and precision of radiotherapy plan formulation. This study also lays the groundwork for the application of more accurate methods in the parameter estimation of NTCP models in the future.

Abbreviations

NTCP	Normal Tissue Complication Probability
IMRT	Intensity-Modulated Radiotherapy
BE	Bayesian estimation
MLE	Maximum Likelihood Estimation
LSE	Least Squares Estimation
LKB	Lyman-Kutcher-Burman
SRU	Serial Reconstruction Unit
AUC	Area Under the Curve
DVH	Different dose-volume histogram
ROC	Receiver Operating Characteristic
AIC	Akaike Information Criterion

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-025-13906-6>.

Supplementary Material 1.

Acknowledgements

Not applicable.

Authors' contributions

L.Z, Z.X and Q.L. conceived of the study, participated in its design and coordination and revised the manuscript. H.O.Y and Y.L. performed the statistical analysis and wrote the original draft. Q.L, X.H, J.Z, L.T, M.L, J.D, R.H, J.H, Z.H, S.D and J.W participated in the data collection, All authors read and approved the final manuscript.

Funding

This study was supported financially, in part, by grant from the Jiangxi natural Science Foundation of China (20202BAB206057), the Applied Research Cultivation Program of Jiangxi Province (20212BAG70047), Beijing Xisike Clinical Oncology Research Foundation (Y-XD202001/zb-0002) and the Second Affiliated Hospital Of NanChang University Funding Program (2022efyB05).

Data availability

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

Declarations

Ethics approval and consent to participate

The study was approved by the Biomedical Research Ethics Committee of the Second Affiliated Hospital of Nanchang University (No.O-2023107).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Vocational Teachers, Jiangxi Agricultural University, Nanchang, Jiangxi Province 330045, P.R. China. ²College of Chemistry and Material, Jiangxi Agricultural University, Nanchang, Jiangxi Province 330045, P.R. China. ³Department of Oncology, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi Province 330006, P.R. China. ⁴College of Computer Science, Sichuan University, Chengdu, Sichuan Province 610065, P.R. China. ⁵The Rehabilitation College of Nanchang University, Nanchang, Jiangxi Province 330031, P.R. China.

Received: 17 November 2024 Accepted: 10 March 2025

Published online: 29 March 2025

References

1. Momeni N, Ali Boroomand M, Roozmand Z, Namiranian N, Hamzian N. Normal tissue complication probability of acute eyelids erythema following radiotherapy of head and neck cancers and skull-base tumors. *Phys Med.* 2023;112:102621.
2. Lai TY, Hu YW, Wang TH, Chen JP, Shiao CY, Huang PI, et al. Estimating the risk of major adverse cardiac events following radiotherapy for left breast cancer using a modified generalized Lyman normal-tissue complication probability model. *Breast.* 2024;77:103788.
3. Rancati T, Fiorino C, Fellin G, Vavassori V, Cagna E, Casanova Borca V, et al. Inclusion of clinical risk factors into NTCP modelling of late rectal toxicity after high dose radiotherapy for prostate cancer. *Radiother Oncol.* 2011;100:124–30.

4. Wang D, Yin Y, Zhou Q, Li Z, Ma X, Yin Y, et al. Dosimetric predictors and Lyman normal tissue complication probability model of hematological toxicity in cervical cancer patients with treated with pelvic irradiation. *Med Phys*. 2022;49:756–67.
5. Jackson A, Ten Haken RK, Robertson JM, Kessler ML, Kutcher GJ, Lawrence TS. Analysis of clinical complication data for radiation hepatitis using a parallel architecture model. *Int J Radiat Oncol Biol Phys*. 1995;31:883–91.
6. Van Dijk LV, Abusaif AA, Rigert J, Naser MA, Hutcheson KA, Lai SY, et al. Normal tissue complication probability (NTCP) prediction model for osteoradionecrosis of the mandible in patients with head and neck cancer after radiation therapy: large-scale observational cohort. *Int J Radiat Oncol Biol Phys*. 2021;111:549–58.
7. Gloi A, McCourt S, Buchanan R, Goettler A, Zuge C, Balzoa P, et al. Dosimetric parameters in partial breast irradiation through brachytherapy. *Med Dosim*. 2009;34:207–13.
8. Tanikawa K, Matsumoto Y, Matsuzaki T, Shimizu M, Matsumoto M, Fukuoka M. A computer program for pharmacokinetics based on maximum likelihood estimation using the gamma distribution with a probability density function: comparison with the normal distribution. *Biol Pharm Bull*. 2000;23:235–9.
9. Tong X, Bentler PM. Evaluation of a new mean scaled and moment adjusted test statistic for sem. *Struct Equ Modeling*. 2013;20:148–56.
10. Söhn M, Yan D, Liang J, Meldolesi E, Vargas C, Alber M. Incidence of late rectal bleeding in high-dose conformal radiotherapy of prostate cancer using equivalent uniform dose-based and dose-volume-based normal tissue complication probability models. *Int J Radiat Oncol Biol Phys*. 2007;67:1066–73.
11. Tucker SL, Cheung R, Dong L, Liu HH, Thames HD, Huang EH, et al. Dose-volume response analyses of late rectal bleeding after radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys*. 2004;59:353–65.
12. Semenenko VA, Li XA. Lyman-Kutcher-Burman NTCP model parameters for radiation pneumonitis and xerostomia based on combined analysis of published clinical data. *Phys Med Biol*. 2008;53:737–55.
13. Momeni N, Broomand MA, Roozmand Z, Hamzian N. Estimating the dose-response relationship for ocular pain after radiotherapy of head and neck cancers and skull base tumors based on the LKB radiobiological model. *J Biomed Phys Eng*. 2023;13:411–20.
14. D'Avino V, Palma G, Liuzzi R, Conson M, Doria F, Salvatore M, et al. Prediction of gastrointestinal toxicity after external beam radiotherapy for localized prostate cancer. *Radiat Oncol*. 2015;10:80.
15. Chapet O, Kong FM, Lee JS, Hayman JA, Ten Haken RK. Normal tissue complication probability modeling for acute esophagitis in patients treated with conformal radiation therapy for non-small cell lung cancer. *Radiother Oncol*. 2005;77:176–81.
16. Zeng L, Huang SM, Tian YM, Sun XM, Han F, Lu TX, et al. Normal tissue complication probability model for radiation-induced temporal lobe injury after intensity-modulated radiation therapy for nasopharyngeal carcinoma. *Radiology*. 2015;276:243–9.

Publisher's Note

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