Journal of Radiation Research, Vol. 59, No. S1, 2018, p. i50–i57
doi: 10.1093/jrr/rry018
Advance Access Publication: 12 March 2018
Special Issue – Approaches to Select Advanced External Radiotherapy

Assessing the uncertainty in a normal tissue complication probability difference (ΔNTCP): radiation-induced liver disease (RILD) in liver tumour patients treated with proton vs X-ray therapy

Keiji Kobashi[1,*,†], Anussara Prayongrat[2,†], Takuya Kimoto[3], Chie Toramatsu[4], Yasuhiro Dekura[2], Norio Katoh[5], Shinichi Shimizu[6,7], Yoichi M. Ito[8] and Hiroki Shirato[7,9]

[1]Department of Medical Physics, Hokkaido University Hospital, North-15 West-7, Kita-ku, Sapporo, 0608638, Japan
[2]Department of Radiation Oncology, Graduate School of Medicine, Hokkaido University, North-15 West-7, Kita-ku, Sapporo, 0608638, Japan
[3]Department of Radiology, Kyoto Prefectural University of Medicine, 465 Kajii-cho, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto, 6028566, Japan
[4]Department of Radiation Oncology, Tokyo Women's Medical University, 8–1,Kawada-cho,Sinjuku,Tokyo, 1628666, Japan
[5]Department of Radiation Oncology, Hokkaido University Hospital, North-15 West-7, Kita-ku, Sapporo, 0608638, Japan
[6]Department of Radiation Oncology, Faculty of Medicine, Hokkaido University, North-15 West-7, Kita-ku, Sapporo, 0608638, Japan
[7]Global Station for Quantum Biomedical Science and Engineering, Global Institute for Cooperative Research and Education, Hokkaido University, North-15 West-7, Kita-ku, Sapporo, 0608638, Japan
[8]Department of Biostatistics, Faculty of Medicine, Hokkaido University, North-15 West-7, Kita-ku, Sapporo, 0608638, Japan
[9]Department of Radiation Medicine, Faculty of Medicine, Hokkaido University, North-15 West-7, Kita-ku, Sapporo, 0608638, Japan
*Corresponding author. Department of Medical Physics, Hokkaido University Hospital, North-15 West-7, Kita-ku, Sapporo, 0608638, Japan.
Tel: +81-11-706-5977; Fax: +81-11-706-7876; Email: kkobashi@med.hokudai.ac.jp
[†]These authors contributed equally to this work.
(Received 23 October 2017; revised 19 February 2018; editorial decision 20 February 2018)

ABSTRACT

Modern radiotherapy technologies such as proton beam therapy (PBT) permit dose escalation to the tumour and minimize unnecessary doses to normal tissues. To achieve appropriate patient selection for PBT, a normal tissue complication probability (NTCP) model can be applied to estimate the risk of treatment-related toxicity relative to X-ray therapy (XRT). A methodology for estimating the difference in NTCP (ΔNTCP), including its uncertainty as a function of dose to normal tissue, is described in this study using the Delta method, a statistical method for evaluating the variance of functions, considering the variance–covariance matrix. We used a virtual individual patient dataset of radiation-induced liver disease (RILD) in liver tumour patients who were treated with XRT as a study model. As an alternative option for individual patient data, dose-bin data, which consists of the number of patients who developed toxicity in each dose level/bin and the total number of patients in that dose level/bin, are useful for multi-institutional data sharing. It provides comparable accuracy with individual patient data when using the Delta method. With reliable NTCP models, the ΔNTCP with uncertainty might potentially guide the use of PBT; however, clinical validation and a cost-effectiveness study are needed to determine the appropriate ΔNTCP threshold.

Keywords: normal tissue complication probability; uncertainty; proton therapy; radiation-induced liver disease; liver tumour

INTRODUCTION

Proton beam therapy (PBT) has emerged as a promising radiotherapy modality due to its favourable physical properties, in that it allows dose escalation to the tumour and minimizes unnecessary doses to normal tissues. With the aim of appropriate patient selection for PBT, a group of researchers at the University of Groningen, Netherlands, first introduced the model-based approach (MBA) to identify patients who could potentially benefit from PBT over X-ray therapy (XRT) by using a normal tissue complication probability (NTCP) model to estimate the risk of developing toxicity [1]. Langendijk *et al.* and Jokabi *et al.* investigated the NTCP difference (ΔNTCP) obtained from *in silico* planning, comparing intensity-modulated radiotherapy (IMRT) with intensity-modulated proton therapy (IMPT) in head and neck cancers [1, 2]. The selective use of PBT relies on a predefined ΔNTCP threshold, such as 10% for Grade II or 15% for total clinical benefit, according to the Dutch Society of Radiation Oncology (NRVO) guidelines [3]. However, these models were derived from a statistical assumption based on a small subset of the population, and the model uncertainty affects the accuracy of PBT selection [4]. In other words, underestimating ΔNTCP can eliminate the opportunity to benefit from PBT, whereas an overly cautious practice might cause the unnecessary use of this high-cost treatment. Therefore, uncertainty regarding the ΔNTCP values is necessary for better clinical implementation in the general population.

In the development of a reliable and generalized NTCP model, using large patient cohorts from multi-institutional datasets can further enhance the model accuracy. However, access to individual dose–volume histogram (DVH) data might be limited in local institutions due to ethical or technical issues. Recently, Wedenberg reported the use of dose-bin data obtained from the literature in order to assess the uncertainty in the estimated dose–response relation for radiation myelopathy and pneumonitis using statistical bootstrap analysis [5]. The advantage of the dose-bin method is its practical convenience, because the data needed for analysis are the number of patients and the number of occurrences of the endpoint in each dose bin. This results in less ethical concern because patient identifier data will not necessarily be revealed. Therefore, dose-bin data is another option for acquiring individual data for large-volume data sharing among institutions. However, the accuracy of using this type of data has yet to be warranted.

The present study uses radiation-induced liver disease (RILD) in liver tumour patients as a study model for describing the methodology of assessing ΔNTCP with uncertainty, using the 95% confidence interval (95% CI), between XRT and PBT. It also compares the results derived from individual patient datasets with those derived from dose-bin datasets, using various methods.

MATERIALS AND METHODS

NTCP model for RILD

RILD is a dose-limiting toxicity of liver radiotherapy and occurs with a frequency of ~5–10% when whole liver is irradiated with up to 30–35 Gy [6, 7], but a tumour requires 60–70 Gy for curative purposes. The most commonly used NTCP model for RILD is the Lyman–Kutcher–Burman (LKB) model [8, 9]. The three parameters of the LKB NTCP model are $TD_{50}(1)$ (the 50% tolerance dose of whole organ), m (the steepness of the dose–response curve) and n (the volume effect). To account for non-homogenous irradiation to the organ, the generalized equivalent uniform dose (gEUD) was adopted [10, 11]. The physical dose from the PBT plan, assuming a RBE of 1.1, and the fractionation schemes should be converted into gEUD as described in a supplementary document (Appendix 1).

Study scheme and virtual patient dataset generation

The study scheme is illustrated in Fig. 1. Due to the lack of a large set of DVH data and observed toxicities in liver cancer patients, we created a virtual patient dataset mimicking the Michigan data [12, 13]. In Monte Carlo fashion, a set of virtual patients in whom the statistics on mean normal liver dose (MNLD) distribution and RILD events were similar to the Michigan data was generated (REF dataset). Subsequently, the REF dataset was organized by dividing the dose level into equal intervals of 5 Gy (0–5 Gy, 5–10 Gy, 10–15 Gy, and so on) and then counting the number of RILD cases and the total number of patients in each dose level/bin (dose bin, DB, dataset) (Appendix 2).

Determination of LKB NTCP model parameters

The two types of dataset included the input data for determining the LKB NTCP model parameters using maximum likelihood estimation (MLE), where TD_{50} (1) and m were adjusted to maximize the probabilities of predicting complications for those who experienced RILD and of predicting no complications for those who did not [14]. Subsequently, the variance (σ^2 or var) and covariance (cov) of TD_{50} (1) and m were obtained from the observed Fisher Information Matrix (FIM). The approximate 95% CI was evaluated as 1.96 standard deviations of the mean (σ or $\sqrt{var}$), $1.96 \times \sqrt{var}$, according to the central limit theorem.

Definition of the ΔNTCP function

With gEUD from the treatment plan and the estimated LKB NTCP parameters, TD_{50} (1) and m, from MLE, the ΔNTCP between XRT and PBT is given by the following function:

$$f(\mathrm{XRT}, \mathrm{PBT}) = \mathrm{NTCP}_{\mathrm{XRT}}(gEUD_{XRT} \mid TD_{50}(1), m) - \mathrm{NTCP}_{\mathrm{PBT}}(gEUD_{PBT} \mid TD_{50}(1), m),$$

where f is a function of ΔNTCP between XRT and PBT and $\mathrm{gEUD}_{\mathrm{XRT}}$ and $\mathrm{gEUD}_{\mathrm{PBT}}$ denote the normal liver dose for a certain patient for XRT and PBT, respectively. TD_{50} (1) is the 50% tolerance dose for uniform distribution for the whole organ, and m is the steepness of the dose–response curve at TD_{50} (1).

Determination of the ΔNTCP with uncertainty

Given the function of the ΔNTCP and the estimated var and cov matrix, the Delta method was applied [15]. Briefly, the Delta method is a standard statistical method for obtaining an approximation of the variance of a function.

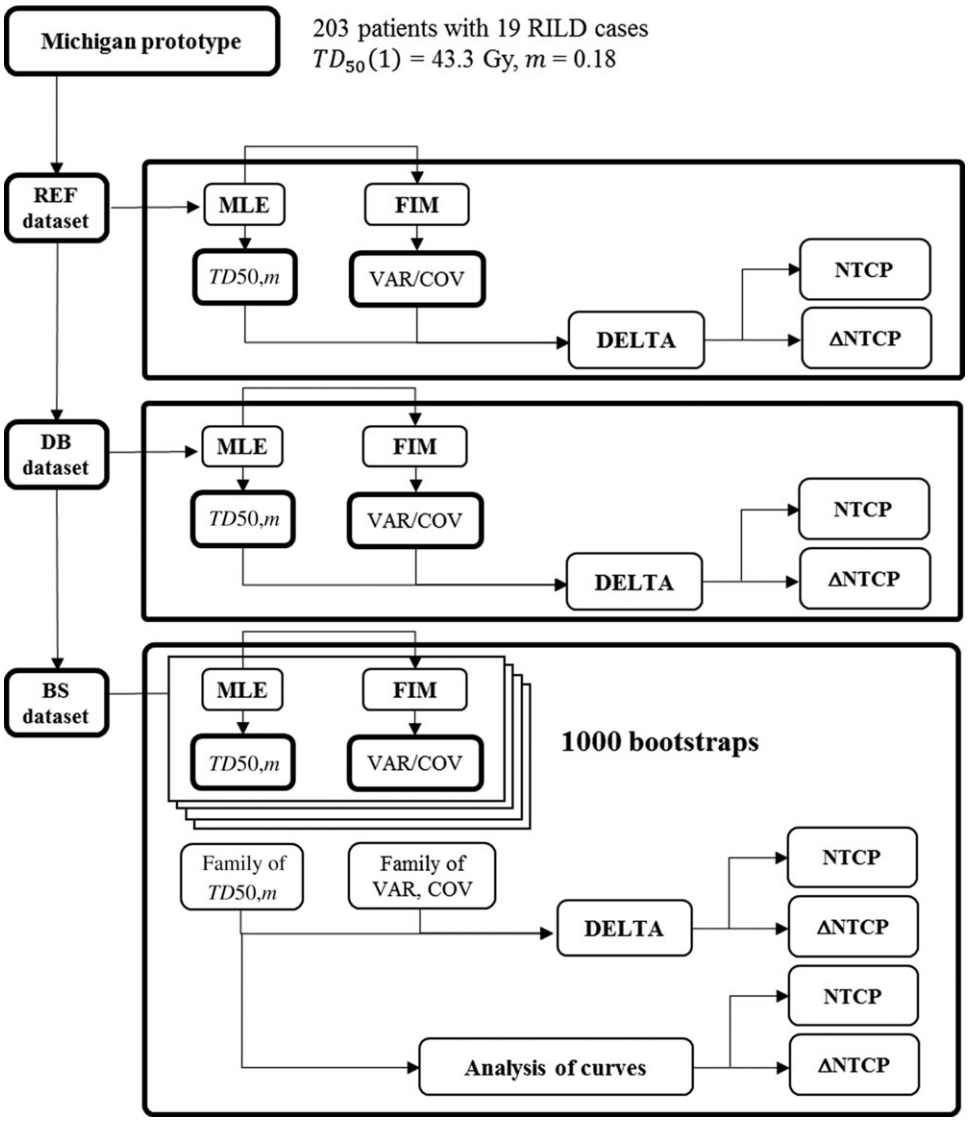

Fig. 1. A study scheme illustrating the proposed algorithms for identifying the NTCP and its difference (ΔNTCP) with 95% confidence intervals. RILD = radiation-induced liver disease, NTCP = normal tissue complication probability, ΔNTCP = NTCP difference, REF = reference, DB = dose-bin, BS = bootstrap, MLE = maximal likelihood estimation, FIM = Fisher information matrix, DELTA = Delta method, TD_{50} (1) = the 50% tolerance dose for whole organ, m = steepness of dose–response curve at $TD_{50}(1)$, VAR = variance, COV = covariance.

$$var(f) = (\partial f/\partial TD_{50})^2 var(TD_{50}) + 2(\partial f/\partial \mathrm{TD}_{50})(\partial f/\partial m)\, cov(TD_{50}, m) + (\partial f/\partial m)^2 var(m)$$

where $\partial f/\partial TD_{50}$ and $\partial f/\partial m$ are partial differentials of f with central estimates of TD_{50} and m, $var(TD_{50})$ and $var(m)$ are variances of TD_{50} and m, respectively, and $cov(TD_{50}, m)$ is a covariance between TD_{50} and m.

Four different algorithms for assessment of the ΔNTCP uncertainty (95% CI) were proposed in the present study. Algorithm #1: the Delta method was applied using the REF dataset to define ΔNTCP with a 95% CI ($\Delta NTCP_{REF}$) as described above. Algorithm #2: the same procedures were performed using the DB dataset ($\Delta NTCP_{DB}$).

Using the bootstrapping technique, newly synthesized individual datasets were generated from the DB dataset. In this case, a thousand bootstrap replicates were generated by random sampling with replacement within each dose-bin. As a result, each bootstrap replicate represented an alternative outcome in a different set of patients of the same size from the same population and contributed to different values of $TD_{50}(1)$ and m and to their variability (var and cov). Algorithm #3: the family of parameters and variability were analysed using a method introduced by Efron [16] to identify the means of $TD_{50}(1)$,m, var and cov. Subsequently, the ΔNTCP with uncertainty

was evaluated using the Delta method ($\Delta NTCP_{BS1}$). Another method for assessing the uncertainty due to sampling variability was proposed by Wedenberg [5] and was used here. Algorithm #4: the 95% bootstrap CI of the ΔNTCP was estimated by analysis of the family of NTCP and ΔNTCP values ($\Delta NTCP_{BS2}$).

All analyses were conducted in R statistics (R Development Core Team, 2010) [17].

RESULTS

Estimated LKB NTCP parameters

From the generated REF dataset (Algorithm #1), the average MNLD among 203 patients was 29.9 Gy (range, 15.2–43.7 Gy), compared with 32 Gy (range, 14.9–44 Gy) in the Michigan study's original data. The average MNLD was 40 Gy and 28.9 Gy for RILD and non-RILD patients, respectively, in the REF dataset and was 37 Gy and 31.3 Gy in the Michigan study. The re-estimated $TD_{50}(1)$ and the parameter m were 43.2 Gy (95% CI 39.1–47.3) and 0.18 (95% CI 0.11–0.24), respectively, which were convincingly similar to the original parameters, 43.3 Gy (95% CI 41.9–52.8) and 0.18 (95% CI 0.14–0.24) [12]. Figure 2 illustrates the estimated NTCP and gEUD of 203 patients with 19 RILD in the REF dataset resembling the Michigan study data (Fig. 2a), with an estimated NTCP curve with 95% CI as a function of gEUD of normal liver, considering $cov(TD50, m)$ (Fig. 2b).

The re-estimated parameters from the DB dataset without (Algorithm #2) and with the bootstrapping technique (Algorithms #3 and 4) were compared with those from the Michigan study and the REF dataset, as shown in Table 1. Comparisons of the NTCP curves of Algorithm #1, #2 and #3 showed nearly identical results (Fig. 2c). Note that no representative NTCP curves, $TD_{50}(1)$ or m were identified for Algorithm #4 due to the nature of the uncertainty estimation by curve analysis.

Estimated ΔNTCP with 95% CI

From the REF dataset (Algorithm #1), the *var* of $TD_{50}(1)$ and m were 4.41 and 0.0010, respectively, which was similar to 4.323 and

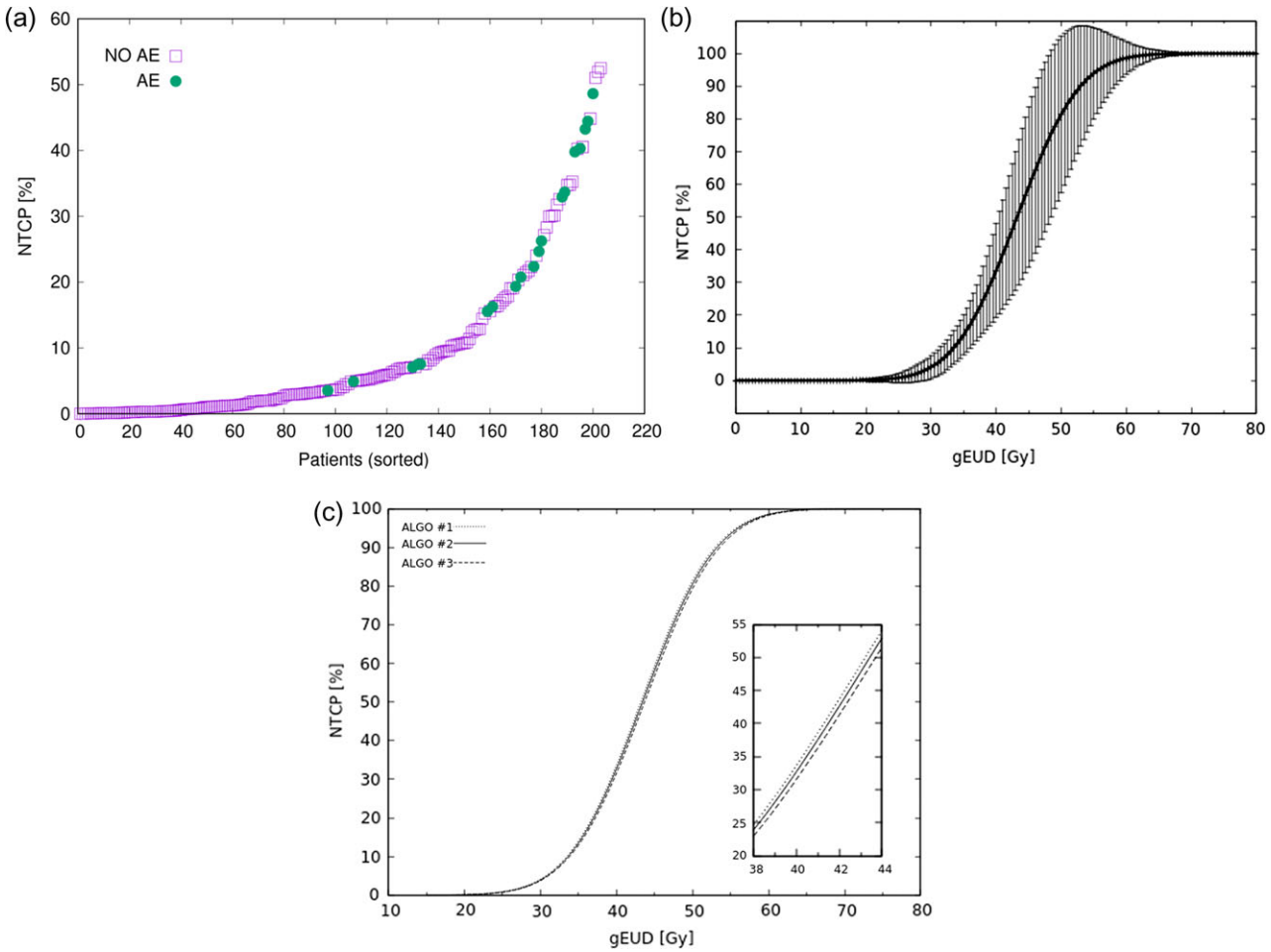

Fig. 2. The generated virtual individual patient dataset (REF dataset) consisting of 203 patients with 19 RILD, mimicking the Michigan study dataset (Fig. 2a), contributed to the estimated NTCP curve with a 95% CI, considering the covariance (Fig. 2b). Comparisons of the NTCP curves from Algorithms #1, #2 and #3 show identical results (Fig. 2c). NTCP = normal tissue complication probability, AE = adverse event, gEUD = generalized equivalent uniform dose, ALGO = algorithm.

Table 1. LKB NTCP parameters with 95% confidence intervals and their variance and covariance, according to the four proposed algorithms

Algorithms		#1	#2	#3	#4[a]
Dataset	Michigan's data	REF dataset	DB dataset	BS from DB dataset	
LKB NTCP parameters (95% CI)					
$TD_{50}(1)$, Gy	43.3 (42.9–52.8)	43.2 (39.1–47.3)	43.4 (39.4–47.5)	43.7 (38.7–48.7)	NA
m	0.18 (0.14–0.24)	0.18 (0.11–0.24)	0.18 (0.11–0.24)	0.18 (0.11–0.24)	NA
Variance					
$TD_{50}(1)$	NA	4.41	4.32	6.47	NA
m	NA	0.0010	0.0010	0.0011	NA
Covariance	NA	0.0519	0.0502	0.0606	NA

[a]No representative NTCP curves, $TD_{50}(1)$, m including their variability were identified for Algorithm #4 due to the nature of uncertainty estimation by curve analysis. LKB = Lyman–Kutcher–Burman, NTCP = normal tissue complication probability, CI = confidence interval, REF = reference, DB = dose-bin, BS = bootstrapping, $TD_{50}(1)$ = the 50% tolerance dose for whole organ, m = steepness of the dose–response curve at $TD_{50}(1)$, NA = not available.

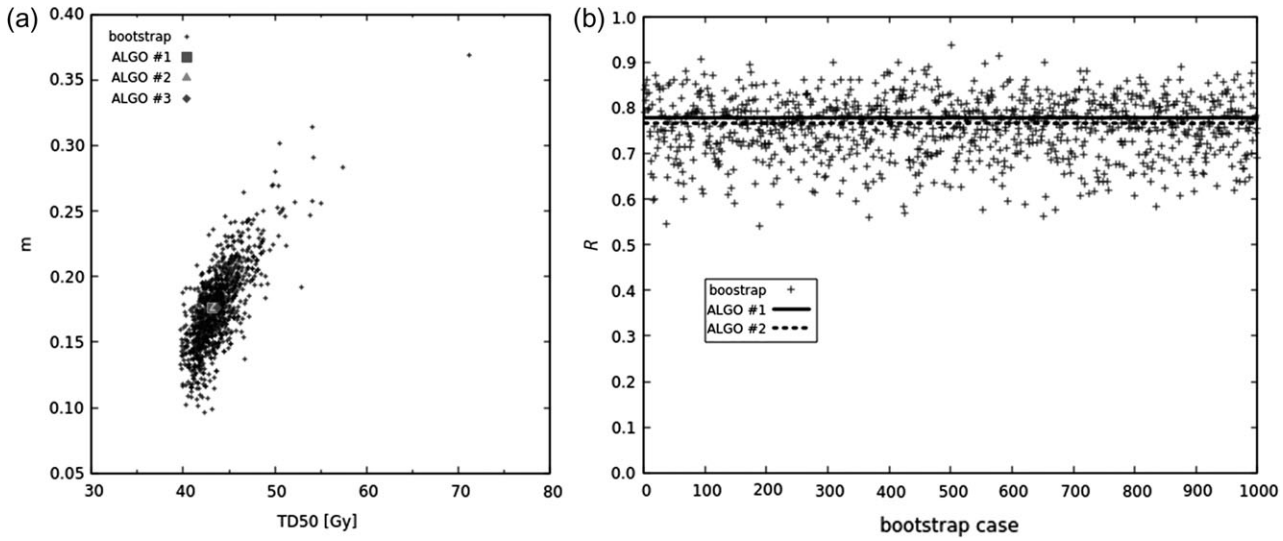

Fig. 3. The distribution of model parameters from Algorithm #1 (■), #2 (▲) and 1000 bootstraps (+) with the mean value #3 (◆) (Fig. 3a). The correlation coefficients (R) between TD_{50} and m from Algorithms #1 (solid line), #2 (dashed line) and #3 1000 bootstrap cases (+) suggested a strong relationship between TD_{50} and m (Fig. 3b). $TD_{50}(1)$ = the 50% tolerance dose for whole organ, m = steepness of dose–response curve at $TD_{50}(1)$, R = Pearson's correlation coefficient, ALGO = algorithm.

0.0010 derived from the DB dataset (Algorithm #2). The $cov(TD50, m)$ was calculated as 0.0519 and 0.502, respectively. The correlation coefficient calculated from *var* and *cov* showed a positive strong relationship between $TD_{50}(1)$ and m, with a value of 0.778 in the REF dataset and 0.768 in the DB dataset. Using the bootstrapping technique, the distribution of the model parameters from 1000 bootstrap replicates is shown in Fig. 3a and b, with a correlation coefficient of 0.8. Again, *var* and *cov* were not assessed in Algorithm #4.

The $\Delta NTCP_{REF}$, as a function of gEUD between XRT and PBT, is shown in the contour lines of the central estimate (Fig. 4a) and the 95% CI lower boundary (Fig. 4b). Each line represents the iso-ΔNTCP. With the contours of ΔNTCP, we can select the patient for PBT based on the confidence level. For example, in Fig. 4c, using the contour of the 95% CI lower boundary, the PBT is favoured for those with a normal liver dose of 36 Gy or more with the XRT plan, with a 95% confidence that the ΔNTCP is 10% or higher. In contrast, using the contours of the central estimate, PBT likely provides a potential benefit for patients who receive a normal liver dose of 33 Gy or more with the XRT plan, with less confidence. Thus, the use of a 95% CI lower boundary contour is more conservative for patient selection for PBT than is using the central estimate contour. However, in the area of 33–36 Gy, the

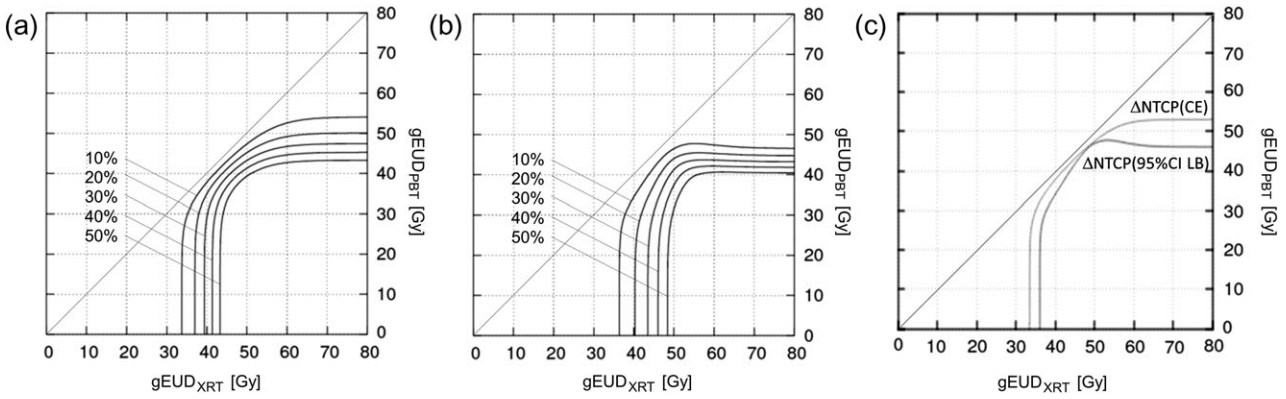

Fig. 4. Using the Delta method in Algorithm #1, the iso-ΔNTCP contours as a function of the gEUD between XRT and PBT are drawn. Each line represents an iso-ΔNTCP with the central estimate (Fig. 4a) and the 95% CI lower boundary (Fig. 4b), varying from 10% to 50%. The 10% ΔNTCP contours of the central estimate (ΔNTCP-CE) and 95% CI lower boundary (ΔNTCP-95% CI LB) are depicted (Fig. 4c). gEUD = generalized equivalent uniform dose; XRT = X-ray therapy; PBT = proton beam therapy; ΔNTCP = normal tissue complication probability difference; CE = central estimate; 95%CI LB = 95% confidence interval lower boundary.

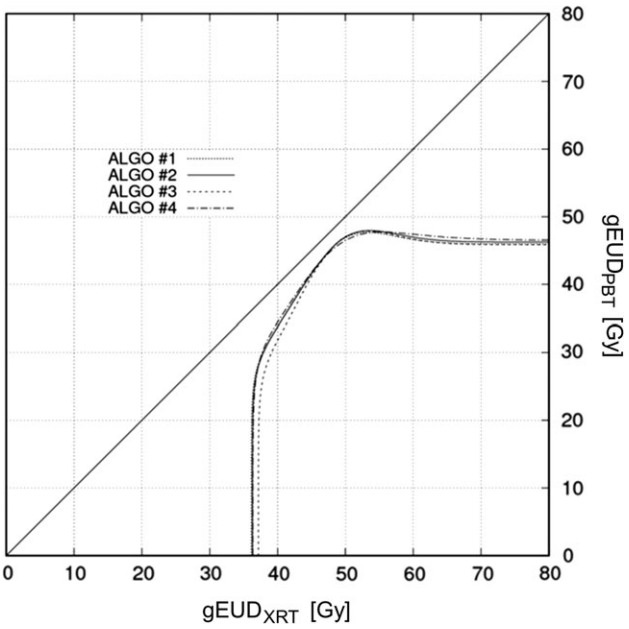

Fig. 5. The contours of the 95% CI lower boundary of 10% ΔNTCP between XRT and PBT derived from four proposed algorithms. gEUD = generalized equivalent uniform dose, XRT = X-ray therapy, PBT = proton beam therapy, ALGO = algorithm.

selection of the treatment modality should be determined by considering the trade-offs between clinical benefits and socio-economic aspects.

The contours of the 95% CI lower boundary of the ΔNTCP at 10% among four proposed algorithms are compared in Fig. 5. Compared with the $\Delta NTCP_{REF}$ (Algorithm #1), the contour of the $\Delta NTCP_{DB}$ (Algorithm #2) was nearly indistinguishable, as were the $\Delta NTCP_{BS1}$ (Algorithm #3) and the $\Delta NTCP_{BS2}$ (Algorithm #4). This was true, in spite of the varying percentage differences of the ΔNTCP. However, the contours from the bootstrapping techniques yielded little differences. This might be due to the larger *cov* found in the bootstrap dataset.

DISCUSSION

Considering the Michigan study's parameters and the number of patients reported in the literature [12], the REF dataset was a good representation of the Michigan data, despite a small difference in the MNLD. The re-identified model parameters were remarkably similar to the original ones. However, variant 95% CI values were observed due to the different methods used to obtain the uncertainty, i.e., the profile likelihood method used in the Michigan study versus the variance-based method based on the central limit theorem used in our study. With virtual simulated gEUD data from 203 patients, the ΔNTCP with 95% CI between the treatment modalities could be determined using the Delta method. The 95% CI lower boundary of the ΔNTCP provides a more conservative threshold for selecting patients for PBT compared with the central estimate. Our results also demonstrated that the LKB NTCP model parameters and variability derived from the dose-bin dataset were very similar to those from the individual patient dataset. Additionally, the contour of the $\Delta NTCP_{DB}$ was very similar to the one of the $\Delta NTCP_{REF}$.

Liver cancer is common in Eastern and Southeastern Asia [18]. Data from Japan on comprehensive cancer statistics show that it is the fifth most common cancer, with an estimated incidence of 45 100 cases, and liver cancer was the fifth most common cause of cancer deaths in 2015[19]. PBT is an effective treatment modality for both primary and secondary liver malignancies [20–25]. A prospective Phase II clinical study demonstrated the efficacy and feasibility of PBT in primary liver tumours [26]. Although a randomized controlled trial (RCT) comparing PBT with standard XRT is expected in the future, there are several obstacles impeding the

conduct of an RCT in this situation [27, 28]. Thus, the MBA-based NTCP model is currently more appealing due to its feasibility in development and implementation for allocating the best treatment modality to individual patients [1, 2, 29].

Bijman *et al.* assessed the model uncertainty using a probability distribution (mean and CI) of the model coefficients from multivariable NTCP models in head and neck cancers and concluded that the accuracy of the MBA on patient selection for PBT is largely affected by the uncertainty in the NTCP models [4]. In contrast, the uncertainty of the LKB NTCP model for RILD relies on the model parameters' variability, *var* and *cov*. Due to the high correlation between TD_{50} and m, we assumed that the uncertainty of the NTCP curves would be better estimated when $cov(TD_{50}, m)$ was considered.

The mean value of the NTCP was generally applied to obtain the ΔNTCP [1–3]. In the present study, we showed that the lower boundary of the 95% CI, considering *var* and *cov*, was a more conservative threshold of ΔNTCP for decision-making regarding the use of PBT. In clinical practice, for a certain patient, a treatment plan comparison between XRT and PBT is performed, based on the dosimetric difference of normal liver DVHs. When using our iso-ΔNTCP curve (Fig. 4), if the gEUD falls within the area where the ΔNTCP is more than the predefined ΔNTCP threshold (the area to the right of and beneath the contour), a PBT can be chosen for a particular patient. This 95% CI lower boundary method conservatively selects those patients who potentially could benefit from PBT. Thus, the overuse of PBT can be prevented. However, the predefined threshold should be appropriately determined based on clinical outcomes and cost-effectiveness studies.

Toramatsu *et al.* performed a dosimetric comparison between spot-scanning proton therapy (SSPT) and IMRT in 10 HCC patients with 13 tumours [30]. We obtained the gEUDs, or mean fraction size equivalent doses (FEDs), from the XRT and PBT plan from table 2 in Toramatsu *et al.*'s publication and applied them to our iso-ΔNTCP contours. According to Toramatsu's study, tumours with a nominal diameter of >6.3 cm (8/13 tumours) had an average risk of RILD of 6.2% for SSPT and 94.5% for IMRT, corresponding with our 10% iso-ΔNTCP 95% CI lower boundary contour (Fig. 6). However, the parameter n was slightly different in our Michigan-resembling NTCP curve ($n = 1.1$) and Toramatsu's publication ($n = 0.97$).

Based on the types of available data, our study suggests that the DB data can be utilized and achieves the same results as individual patient data, with more convenience in data collection and sharing. Algorithm #2 constantly provided results similar to those achieved with Algorithm #1, whereas Algorithms #3 and #4 were affected by statistical instability and consumed a large amount of time.

Due to the lack of individual patient data, we created a virtual simulated patient dataset based on the Michigan study data, in which the fractionation scheme was unique (1.5 Gy twice daily) and concurrent chemotherapy was administered. In this study, the LKB model was used to create the map, but there were several types of NTCP models assuming different equations and parameters and resulting in different NTCP values and ΔNTCP contours. Thus, our iso-ΔNTCP contours should be interpreted cautiously. Furthermore, the NTCP model derived from the Michigan study was dedicated to patients with normal liver function treated with XRT, and

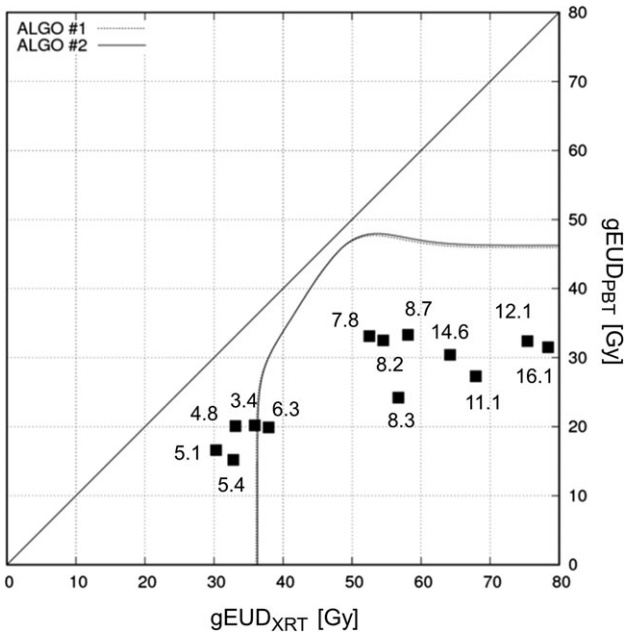

Fig. 6. Fraction size equivalent doses (FEDs) of 13 liver tumours from Toramatsu's data (filled squares) are plotted with the contour of the iso-ΔNTCP 95% CI lower boundary derived from Algorithms #1 (dashed line) and #2 (solid line). The small numbers above denote tumour size in centimeters. gEUD = generalized equivalent uniform dose, XRT = X-ray therapy, PBT = proton beam therapy, ALGO = algorithm.

the PBT is theoretically useful for those with impaired hepatic function (Child–Pugh B and C), as suggested by Dawson [31]. As a result, the NTCP model needs to be prospectively developed and needs to include these patient subgroups in order to develop an accurate and reliable iso-ΔNTCP decision-making map. In the future, clinical validation studies and cost-effectiveness analyses are expected to select patients who potentially could benefit from PBT.

In conclusion, the methodology presented in this paper relies on systematic statistical considerations of the 95% CI based on the Delta method and, considering the variance–covariance matrix, can be applied to other types of NTCP models and tumours. This might ultimately establish a guideline for properly selecting patients for proton therapy.

SUPPLEMENTARY DATA

Supplementary data is available at *Journal of Radiation Research* online.

CONFLICT OF INTEREST

The authors state that there are no conflicts of interest.

REFERENCES

1. Langendijk JA, Lambin P, De Ruysscher D et al. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. *Radiother Oncol* 2013;107:267–73.

2. Jakobi A, Bandurska-Luque A, Stutzer K et al. Identification of patient benefit from proton therapy for advanced head and neck cancer patients based on individual and subgroup normal tissue complication probability analysis. *Int J Radiat Oncol Biol Phys* 2015;92:1165–74.
3. Cheng Q, Roelofs E, Ramaekers BLT et al. Development and evaluation of an online three-level proton vs photon decision support prototype for head and neck cancer—comparison of dose, toxicity and cost-effectiveness. *Radiother Oncol* 2016;118:281–5.
4. Bijman RG, Breedveld S, Arts T et al. Impact of model and dose uncertainty on model-based selection of oropharyngeal cancer patients for proton therapy. *Acta Oncol* 2017:56:1444–50.
5. Wedenberg M. Assessing the uncertainty in QUANTEC's dose–response relation of lung and spinal cord with a bootstrap analysis. *Int J Radiat Oncol Biol Phys* 2013;87:795–801.
6. Emami B, Lyman J, Brown A et al. Tolerance of normal tissue to therapeutic irradiation. *Int J Radiat Oncol Biol Phys* 1991;21:109–22.
7. Benson R, Madan R, Kilambi R et al. Radiation induced liver disease: a clinical update. *J Egypt Natl Canc Inst* 2016;28:7–11.
8. Lyman JT. Complication probability as assessed from dose–volume histograms. *Radiat Res Suppl* 1985;8:S13–9.
9. Kutcher GJ, Burman C, Brewster L et al. Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. *Int J Radiat Oncol Biol Phys* 1991;21:137–46.
10. Luxton G, Keall PJ, King CR. A new formula for normal tissue complication probability (NTCP) as a function of equivalent uniform dose (EUD). *Phys Med Biol* 2008;53:23–6.
11. Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. *Med Phys* 1997;24:103–10.
12. Dawson LA, Normolle D, Balter JM et al. Analysis of radiation-induced liver disease using the Lyman NTCP model. *Int J Radiat Oncol Biol Phys* 2002;53:810–21.
13. Ten Haken RK. WE-C-BRA-04: Optimizing radiotherapy using NTCP models: 17 years in Ann Arbor. Abstract for American Association of Physicists in Medicine (AAPM). *Med Phys* 2007;34:2595. doi:10.1118/1.2761535.
14. Martin C, Brad O, Kerwyn F et al. An MLE method for finding LKB NTCP model parameters using Monte Carlo uncertainty estimates. *J Phys Conf Ser* 2014;489:012087.
15. Armitage P, Berry G, Matthews J. *Statistical Methods in Medical Research*, 4th edn. UK: Blackwell Science; 2001.
16. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman & Hall, 1993.
17. Team RDC. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2010.
18. Ferlay J, Soerjomataram I, Dikshit R et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.
19. Hori M, Matsuda T, Shibata A et al. Cancer incidence and incidence rates in Japan in 2009: a study of 32 population-based cancer registries for the Monitoring of Cancer Incidence in Japan (MCIJ) project. *Jpn J Clin Oncol* 2015;45:884–91.
20. American Medical Association. *ASTRO Model Policies: Proton Beam Therapy*, 2014. https://www.astro.org/uploadedFiles/Main_Site/Practice_Management/Reimbursement/ASTRO%20PBT%20Model%20Policy%20FINAL.pdf (23 October 2017, date last accessed).
21. Tanguturi SK, Wo JY, Zhu AX et al. Radiation therapy for liver tumors: ready for inclusion in guidelines? *Oncologist* 2014;19:868–79.
22. Durante M. Charged particles for liver cancer. *Ann Transl Med* 2015;3:363.
23. Qi WX, Fu S, Zhang Q et al. Charged particle therapy versus photon therapy for patients with hepatocellular carcinoma: a systematic review and meta-analysis. *Radiother Oncol* 2015;114:289–95.
24. Granovetter M. Proton radiotherapy for primary liver cancers. *Lancet Oncol* 2016;17:e49.
25. Durante M, Orecchia R, Loeffler JS. Charged-particle therapy in cancer: clinical uses and future perspectives. *Nat Rev Clin Oncol* 2017;14:483–95.
26. Hong TS, Wo JY, Yeap BY et al. Multi-institutional phase II study of high-dose hypofractionated proton beam therapy in patients with localized, unresectable hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *J Clin Oncol* 2016;34:460–8.
27. Hellman S, Hellman DS. Of mice but not men. Problems of the randomized clinical trial. *N Engl J Med* 1991;324:1585–9.
28. Goitein M, Cox JD. Should randomized clinical trials be required for proton radiotherapy? *J Clin Oncol* 2008;26:175–6.
29. Blanchard P, Wong AJ, Gunn GB et al. Toward a model-based patient selection strategy for proton therapy: external validation of photon-derived normal tissue complication probability models in a head and neck proton therapy cohort. *Radiother Oncol* 2016;121:381–6.
30. Toramatsu C, Katoh N, Shimizu S et al. What is the appropriate size criterion for proton radiotherapy for hepatocellular carcinoma? A dosimetric comparison of spot-scanning proton therapy versus intensity-modulated radiation therapy. *Radiat Oncol* 2013;8:48.
31. Dawson LA. Protons or photons for hepatocellular carcinoma? Let's move forward together. *Int J Radiat Oncol Biol Phys* 2009;74:661–3.