HOKKAIDO UNIVERSITY

Title	Study on selective use of proton therapy in primary liver cancers using Normal Tissue Complication Probability (NTCP) model
Author(s)	プラヨンラット, アヌサラー
Description	配架番号：2492
Degree Grantor	北海道大学
Degree Name	博士(医学)
Dissertation Number	甲第13716号
Issue Date	2019-09-25
DOI	https://doi.org/10.14943/doctoral.k13716
Doc URL	https://hdl.handle.net/2115/93227
Type	doctoral thesis
File Information	Anussara_Prayongrat.pdf, この論文の全文

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学 位 論 文

Studies on selective use of proton therapy in primary liver cancers using Normal Tissue Complication Probability (NTCP) model

(正常組織有害反応発生確率モデルを用いた肝臓がん患者に対する陽子線治療の選択的適用に関する研究)

2019 年 9 月

北 海 道 大 学

アナサラー ・プラヨンラット

Anussara Prayongrat

学 位 論 文

Studies on selective use of proton therapy in primary liver cancers using Normal Tissue Complication Probability (NTCP) model

(正常組織有害反応発生確率モデルを用いた肝臓がん患者に対する陽子線治療の選択的適用に関する研究)

2019 年 9 月

北 海 道 大 学

アナサラー ・プラヨンラット

Anussara Prayongrat

Table of Contents

List of publications and presentations 1

Summary 2

List of abbreviations 6

Introduction 8

Chapter 1

- Introduction 23
- Materials and methods 24
- Results 30
- Discussion 36

Chapter 2

- Introduction 40
- Materials and methods 41
- Results 43
- Discussion 52

Chapter 3

- Introduction 56
- Materials and methods 56
- Results 58
- Discussion 67

Conclusion 70

Acknowledgements 72

References 73

List of publications and presentations

Part of the research has been published as listed below:

1. Kobashi K*, Prayongrat A*, Kimoto T, Toramatsu C, Dekura Y, Katoh N, Shimizu S, Ito YM, Shirato H, 2018, 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. *Journal of Radiation Research*. vol.59, no. S1, p.i50-i57.

*These authors contributed equally to this work.

2. Prayongrat A, Kobashi K, Ito YM, Katoh N, Tamura M, Dekura Y, Toramatsu C, Khorprasert C, Amornwichet A, Alisanant P, Shirato H, Shimizu S, 2019, The normal tissue complication probability model-based approach considering uncertainties for the selective use of radiation modality in primary liver cancer patients. *Radiotherapy and Oncology*. vol.135, p.100-106.

3. Prayongrat A, Ito YM, Katoh N, Tamura M, Dekura Y, Toramatsu C, Khorprasert C, Amornwichet A, Alisanant P, Hirata Y, Shirato H, Shimizu S, Kobashi K. Multivariable normal tissue complication probability (NTCP) model for predicting radiation-induced liver disease in primary liver cancer (submitted to *International Journal of Radiation Oncology, Biology, Physics*).

Part of the research has been presented in the conference as listed below:

1. Prayongrat A, Kobashi K, Ito YM, Katoh N, Dekura Y, Amornwichet N, Shimizu S, Shirato H, Uncertainties of normal tissue complication probability (NTCP) and NTCP difference between radiation treatment modalities for radiation-induced liver toxicity in Child-Pugh A primary liver cancer patients. 60th American Society for Radiation Oncology (ASTRO), October 21-24, 2018, Texas, USA

Summary

Background and purpose:

Radiotherapy has been one of the backbone treatments in primary liver cancers including hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC). But radiation-induced liver disease (RILD) remains a dose-limiting complication of traditional liver-directed radiotherapy and can lead to deterioration of liver function followed by liver failure and death. With many attempts to improve tumor control and minimize toxicity, modern radiotherapy techniques have been emerging but come at high cost. Proton beam therapy (PBT) has both physical and biological advantages and is effective for primary liver cancers. Although randomized controlled trial (RCT) between PBT *versus* standard x-ray treatment (XRT) is expected, there were several obstacles in conducting RCT. Normal tissue complication probability (NTCP) model-based approach is currently captivating as alternative method.

For primary liver cancers, NTCP model translates dosimetry of normal liver into estimated risk of RILD. However, clinical factors also influence the radiosensitivity of liver tissue. Therefore, NTCP model considering both clinical and dosimetric parameters in primary liver cancers needs to be established. Then, for each patient, risk of RILD is estimated from PBT plan ($NTCP_{PBT}$) and XRT plan ($NTCP_{XRT}$) and compared, $NTCP_{XRT}$ - $NTCP_{PBT}$, referred as ΔNTCP. The patient is eligible for PBT if his/her ΔNTCP is larger than predefined ΔNTCP threshold.

Nevertheless, there are considerable sources of model uncertainties. For example, sample data have been used to develop the model. The lack of exact knowledge in whole population results in model uncertainty which affects the accuracy of NTCP model and ΔNTCP values. Estimation of uncertainties of NTCP and ΔNTCP helps improve the reliability of the model prediction in the general population. Underestimation of ΔNTCP can lead to the loss of opportunity to benefit from PBT while overly cautious practice might cause unnecessary use. Therefore, assessing uncertainty is an important step of model-based patient selection strategy for PBT.

The aims of our studies are to establish the NTCP model by integrating clinical and dosimetric factors for better prediction of RILD risk and to demonstrate the NTCP model-based approach considering uncertainties of NTCP and ΔNTCP to improve the clinical implementation of patient selection for PBT.

Materials and Methods:

We studied the methodology to assess uncertainties of NTCP and ΔNTCP considering variance and covariance of the model parameters from Lyman-Kutcher-Burman (LKB) NTCP model. Virtual simulated data of 203 patients were generated mimicking a landmark (Michigan) study and classified into 3 types of dataset (individual patient, organized dose-bin and bootstrap dataset) to estimate for LKB NTCP parameters, their variance and covariance. Different methods were used to obtain confidence interval (CI) of NTCP and ΔNTCP, namely, delta method and bootstrapping method.

Subsequently, we collected data of 239 eligible primary liver cancer patients and modeled dose-response relationship between mean liver dose (MLD) and grade≥2 liver toxicity according to Common Terminology Criteria for Adverse Events (CTCAE grade≥2 RILT) based on LKB NTCP model. Clinical characteristics including Child-Pugh classification and viral hepatitis infection were recorded. We developed subgroup-specific LKB NTCP models and determined model parameters, their variance and covariance. Then, the delta method was used to estimate ΔNTCP with uncertainty.

Lastly, we introduced the multivariable NTCP model for predicting RILD using 280 patients for training dataset and 46 patients for test dataset. The clinical and dosimetric parameters were analyzed using multivariate regression analysis. Final model was reported, and model performance and validation were performed.

Results:

The virtual simulated patient data was a good representative of the Michigan data despite a small difference in MLD. The re-identified model parameters were remarkably comparable to the original ones. The uncertainty of ΔNTCP was determined by delta method. Our results demonstrated that LKB NTCP model parameters, their variance and covariance, and ΔNTCP contours derived from dose-bin dataset were very similar to those from individual patient dataset. However, the contours from bootstrapping techniques yielded a little difference which might be due to larger covariance found in bootstrap dataset. Additionally, we proposed that the use of CI-lower boundary of ΔNTCP provided more conservative threshold compared to central estimate.

From 239 patients dataset, CP classification was the strongest clinical predictive factor for CTCAE grade≥2 RILT, followed by viral hepatitis infection. We stratified patients into 3 subgroups according to CP classification and hepatitis status, i.e., CP-A, nonhepatitis subgroup; CP-A, hepatitis subgroup; and CP-B/C subgroup. Estimated LKB NTCP parameters and NTCP curves were unique for each subgroup. Likewise, ΔNTCP varied between subgroups. Lower boundary of CI of ΔNTCP was more conservative in all subgroups.

Finally, we firstly reported the multivariable NTCP model for predicting RILD using relatively large patient dataset. The model consisted of 4 parameters including diagnosis (HCC *versus* ICC), pretreatment CP classification (CP-A *versus* CP-B/C), viral hepatitis infection status (positive *versus* negative) and MLD. The model had good performance and validity. We then classified patients according to these 3 clinical risks (not included MLD) into 8 subgroups and estimated their baseline NTCP values. Patients with ≥2 clinical risk factors had increased risk of RILD. Uncertainty of model parameters (regression coefficients) were reported for further studies.

Discussion:

To assess uncertainties of LKB NTCP and ΔNTCP, we found that delta method considering variance and covariance of model parameters was effective for both individual dataset and dose-bin dataset whereas bootstrapping method was limited by statistical instability and longer computational time. This suggested the feasibility of using dose-bin dataset for data sharing among institutions.

According to our 239 patients, assuming a different CP classification and hepatitis status, NTCP and ΔNTCP markedly varied indicating different radiosensitivity among patients. This finding suggested the use of subgroup-specific LKB NTCP model and further underlined the differential benefit from PBT between subgroups which may have a potential impact on the decision regarding selective PBT use. In addition, these studies suggested that lower boundary of CI of ΔNTCP provided more conservative threshold and increased confidence for patient selection for PBT.

More clinical and dosimetric parameters were studied using the multivariable analysis. The 4-variable model consisted of 3 clinical risk features and 1 dosimetric parameter. Patients with 2 clinical risk factors had increased baseline risk of RILD. As a result, treatment plan should be

optimized with high respect to MLD based on NTCP model in these patients. For those with all 3 risk factors, radiotherapy might not be an appropriate treatment due to very high baseline NTCP value regardless of MLD. The uncertainty of regression coefficients reported in this study would be further analyzed for obtaining NTCP and ΔNTCP uncertainties in our future studies.

Nevertheless, appropriate ΔNTCP threshold as well as proper confidence level for the selective use of PBT is still unknown. Therefore, challenges are to determine the appropriate threshold for identifying patients who would most likely benefit from PBT based on clinical outcomes and cost-effectiveness study.

Conclusion:

NTCP model for predicting RILD was established based on subgroup-specific LKB model and multivariable model. We proposed a clinical decision support using NTCP model-based approach to select patients for PBT and suggested the use of ΔNTCP with uncertainty to improve confidence in decision making.

List of Abbreviations

AE	adverse event
ALP	alkaline phosphatase
ALT	alamine aminotransferase
AST	aspartate aminotransferase
ASTRO	American Society for Radiation Oncology
AUC	area under the receiver operating characteristic curve
BS	bootstrap
3D-CRT	three-dimension conformal radiotherapy
CE	central estimate
CI	confidence interval
CP	Child-Pugh
CT	computed tomography
CTCAE	common terminology criteria for adverse events
DB	Dose-bin
DVH	dose-volume histogram
ECOG	Eastern Cooperative Oncology Group
EORTC	European Organization for Research and Treatment of Cancer
EUD	equivalent uniform dose
FED	fraction-size equivalent dose
Gy	Gray (unit of radiation dose)
HBV	viral hepatitis B
HCC	hepatocellular carcinoma
HCV	viral hepatitis C
HL	Hosmer-Lemeshow
ICC	intrahepatic cholangiocarcinoma
IMPT	Intensity-modulated proton therapy
IMRT	Intensity-modulated radiotherapy
INR	international normalized ratio
LB	lower boundary
LKB	Lyman-Kutcher-Burman
MeV	million electron volts
MLD	mean liver dose
MLE	maximum likelihood estimation
MRI	magnetic resonance image
NTCP	normal tissue complication probability
ΔNTCP	normal tissue complication probability reduction
OAR	organ-at-risk
OR	odds ratio
OS	overall survival
PBT	proton beam therapy

PVT	portal vein thrombosis
QUANTEC	Quantitative Analyses of Normal Tissue Effects in the Clinic
RBE	relative biologic effectiveness
RCT	Randomized controlled trial
RILD	radiation-induced liver disease
RILT	radiation-induced liver toxicity
SBRT	stereotactic body radiotherapy
SD	standard deviation
SOBP	spread-out Bragg peak
TACE	transarterial chemoembolization
TD	Tolerance dose
ULN	upper limit of normal
V_{eff}	effective volume
VMAT	volumetric modulated arc therapy
Vx	volume of organ receiving ≥x Gy
XRT	x-ray treatment

Introduction

Liver cancer is the fifth most common cancer in males and the ninth in females. The highest incidence was in Eastern Asia with the age-standardized rate of 26.8 and 8.7 per 100,000 in male and female, respectively. In Japan, the projected incidence of liver cancer in 2018 is 39,600 new cases (the seventh most common of all sites), 25,700 males and 13,900 females, with the estimated 27,000 deaths (the fifth most common causes of cancer deaths). In Southeastern Asia including Thailand, the age-standardized incidence rates were 21 and 6.6 per 100,000 in male and female (Bray et al., 2018).

Primary liver cancers consist of hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC). Radiotherapy has been one of the mainstay treatments in primary liver cancers and can be administered as adjuvant treatment after surgery and definite treatment for unresectable diseases, either radiotherapy alone or combination with chemotherapy. Historically, radiotherapy has a limited role for treatment of liver tumors due to low liver tolerance. Emami et al. reported that tolerance dose for whole liver irradiation with the risk of toxicity at 50% (TD_{50}) and 5% (TD_5) were 40 and 30 Gy, respectively (Emami et al., 1991), whereas radiation dose of more than 60-70 Gy is usually required for effective tumor control. Therefore, conventional x-ray treatment (XRT) has not been recognized as a curative treatment for liver cancers because it is difficult to give adequately high dose to the tumor and safely spare surrounding normal tissue at the same time. An important dose-limiting complication of traditional liver radiotherapy is radiation-induced liver disease (RILD).

Radiation-induced liver disease

RILD typically occurs between 2 weeks to 4 months after completion of radiotherapy and can be followed by liver failure and death due to limited treatment options. RILD can be classified into “classic” and “nonclassic” RILD. Clinical manifestation of classic RILD includes anicteric hepatomegaly, ascites, and elevated alkaline phosphatase (ALP) more than twice the upper limit of normal (2xULN) or baseline value. Pathological characteristics are occlusion and obliteration of the central veins of hepatic lobules, retrograde congestion and secondary hepatocyte necrosis (Reed and Cox, 1966). In contrast, nonclassic RILD involves elevated liver transaminases (aspartate aminotransferase, AST, and alanine aminotransferase, ALT) more than 5xULN within

3 months after therapy, or liver function deterioration measured by a decline in Child-Pugh (CP) score by 2 or more (Table 1), with the absence of classic RILD. This endpoint is usually described in patients with pretreatment poor liver function, for example, hepatitis B infection, CP B and C (Pan et al., 2010).

Table 1 Child-Pugh Scoring system

Criteria	1 point	2 points	3 points
Total bilirubin	<2 mg/dL	2-3 mg/dL	>3 mg/dL
Serum albumin	>3.5 g/dL	2.8-3.5 g/dL	<2.8 g/dL
INR	<1.7	1.71-2.20	>2.20
Ascites	None	Controlled with medication	Refractory
Hepatic encephalopathy	None	CTCAE grade I-II (or controlled with medication)	CTCAE grade III-IV (or refractory)
Total score of 5-6 is Child-Pugh class A			
Total score of 7-9 is Child-Pugh class B			
Total score ≥10 is Child-Pugh class C			

Abbreviation: INR=international normalized ratio; CTCAE=common terminology criteria for adverse events.

RILD is seen in 5%-10% of patients who received 30-50 Gy to whole liver whereas tumor needs about 60-70 Gy for curative aim as mentioned above. Therefore, unfavorable local tumor control results from inadequate radiation dose from conventional technique. Nonetheless, with the advances in radiotherapy technology, modern radiotherapy techniques such as stereotactic body radiotherapy (SBRT) and charge particle therapy including proton beam therapy (PBT) allows dose escalation to the tumor with accurate beam delivery. As a result, these cutting-edge technologies could provide curative goal with minimal toxicity (Tanguturi et al., 2014, Tao and Yang, 2012).

Risk factors for RILD can be classified into 3 categories:

(1) patient-related risk factors

Male gender, CP classification and viral hepatitis infection status were associated with higher risk of RILD. A study showed a significantly increased risk of RILD in male gender with Odds ratio (OR) of 9.9 (*p-value=0.002*) (Dawson et al., 2002). Several studies reported the greater susceptibility to RILD in patients with CP-B, and viral hepatitis B (HBV) carrier status with OR of 3.7-13.6 for CP-B (*p-value<0.001*) and OR of 9.3 for HBV carrier (*p-value=0.040*) (Liang et al., 2006, Xu et al., 2006, Cheng et al., 2004, Prayongrat A et al., 2019).

(2) disease-related risk factors

Diagnosis of primary liver cancers (*versus* liver metastases), presence of portal vein thrombosis (PVT), tumor characteristics defined by cancer staging as well as tumor size were studied. Patients with primary liver cancers diagnosis had significantly higher risk of RILD with OR of 6.4 (*p-value=0.005*) for hyperfractionated radiotherapy (Dawson et al., 2002, Osmundson et al., 2015) and OR of 5.8 (*p-value=0.009*) for SBRT (Osmundson et al., 2015). The presence of PVT, late tumor stage and large gross tumor volume were only associated with RILD in univariate analysis but not in multivariate analysis (Liang et al., 2006, Seong et al., 2003). However, PVT was found to be significantly associated with grade ≥2 liver toxicity according to Common Terminology Criteria for Adverse Events (CTCAE) (Prayongrat A et al., 2019).

(3) treatment-related risk factors

Various dosimetric parameters associated with RILD were mean normal liver dose, volume of liver receiving 5-30 Gy (V5-30) and normal/total liver volume (Osmundson, 2015). Other factors included chemotherapy and transarterial chemoembolization (TACE) (Dawson et al., 2002, Liang et al., 2006). Administration of continuous infusion hepatic artery bromodeoxyuridine chemotherapy concurrently with twice-daily radiotherapy was significantly associated with RILD with very high OR of 71 (*p-value<0.001*) (Dawson et al., 2002) while TACE was only significant in univariate analysis (Liang et al., 2006).

Proton beam therapy in liver cancers

Delivery of curable dose to the tumor while sparing surrounding normal tissue is very challenging for liver tumors. Charged particles including proton offers physical benefits over conventional X-ray by virtue of Bragg peak. In PBT, proton particles are accelerated to a high speed, up to two-thirds the speed of light, and energies, up to 250 million electron volts (MeV), by an accelerator (synchrotron or cyclotron). The accelerated protons are then steered by magnets toward the treatment room and to a patient through a beam delivery system. These high-energy protons can penetrate the patient's body and stop at a certain range according to their energy. At the end of its range, the dose delivered to tissue is maximized, so-called Bragg peak, followed by a steep gradient with no exit dose because very few protons penetrate beyond that distance.

In cancer treatment, protons of different energies with several Bragg peaks at different depths are used to treat the entire tumor as shown in Figure 1. Along beam entrance path, tissues close to the surface of the body receive lower radiation dose compared to XRT. Similarly, tissues beyond the tumor receive minimal dose along beam exit path. Therefore, unnecessary radiation dose to the surrounding normal tissues is markedly reduced thus allows dose escalation to the tumor which potentially translates into better clinical outcomes.

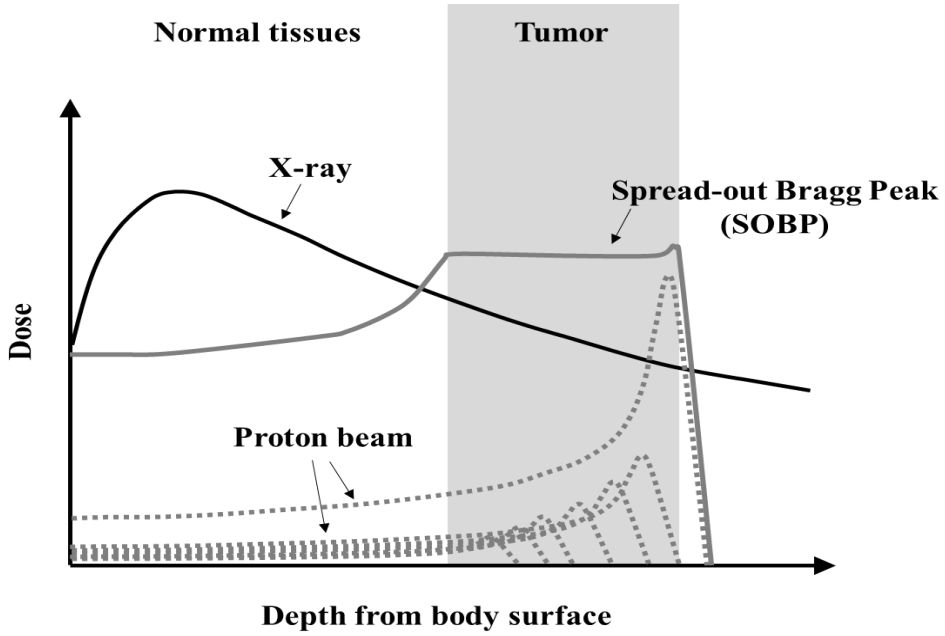


Figure 1 Comparison of x-ray depth dose curve (black) with the Bragg peak (gray) and spread-out Bragg peak (SOBP) (dashed grey) of proton beam therapy

According to the American Society for Radiation Oncology (ASTRO) Proton Beam Therapy Model Policy 2017, HCC is classified as Group 1 (medically necessary) disease site that frequently supports the use of PBT. Studies of PBT in treating patients with HCC are shown in Table 2. PBT was usually administered in hypofractionation scheme and results in favorable outcomes with local control and overall survival (OS) of approximately 80% and 40% at 5 years, respectively.

Table 2 Studies of proton beam therapy in treating patients with hepatocellular carcinoma.

Study	N	CP	Viral hepatitis	Tumor (median) size/ location	PBT dose (GyRBE/ fractions)	Local control	Progression-free survival	Overall survival	Grade≥2 toxicity
(Kawashima et al., 2005)	30	A: 67% B: 33%	HBV: 10% HCV: 87% both: 3%	4.5 cm	76/20	2Y: 96%	2Y: 38% 3Y: 16%	2Y: 66% 3Y: 62%	Gr 3: 33% (hemato-toxicity)
(Chiba et al., 2005)	162	A: 51% B: 38% C: 6%	HBV: 9% HCV: 80% both: 2%	3.8 cm	72/16	5Y: 87%	N/A	5Y: 24%	Gr≥2: 3%
(Mizumoto et al., 2008)	53	A: 87% B: 11% C: 2%	HBV: 9% HCV: 68% both: 2%	4.3 cm/ adjacent to porta hepatis	72.6/22	2Y: 94% 3Y: 86%	2Y: 38% 3Y: 25%	2Y: 57% 3Y: 45%	Gr2: 6%
(Fukumitsu et al., 2009)	51	A: 80% B: 20%	HBV: 16% HCV: 63% both: 4%	2.8 cm/ >2cm away from GI tract	66/10	3Y: 95% 5Y: 88%	N/A	3Y: 49% 5Y: 39%	Gr2: 6%
(Nakayama et al., 2009)	318	A: 74% B: 24% C: 2%	HBV: 7% HCV: 77% both: 1%	N/A	66/10 72.6/22 77/35	N/A	N/A	1Y: 90% 3Y: 65% 5Y: 45%	Gr2: 10% Gr3: 2%

Study	N	CP class	Viral heptatitis	Tumor (median) size/ location	PBT dose (GyRBE/ fractions)	Local control	Progression-free survival	Overall survival	Grade≥2 toxicity
(Nakayama et al., 2011)	47	A: 75% B: 19% C: 6%	HBV: 13% HCV: 72% both: 2%	N/A / within 2 cm away from GI tract	72.6/22 77/35	3Y: 88% 4Y: 88%	3Y: 40% 4Y: 17%	3Y: 50% 4Y: 34%	Gr2: 3% Gr3: 2%
(Bush et al., 2011)	76	A: 29% B: 47% C: 24%	N/A	5.5 cm	63/15	80%	3Y: 60%	3Y: 70% if liver transplant	Gr2: 7%
(Bush et al., 2016)	33	N/A	HBV: 3% HCV: 67%	3.2 cm	70.2/15	2Y: 88%	2Y: 48%	2Y: 59%	No severe toxicity
(Hong et al., 2016)	44	A: 73% B: 21%	HBV: 9% HCV: 52%	5 cm	67.5/15 58.05/15	2Y: 95%	2Y: 40%	2Y: 63%	Gr 3: 5%
(Kim et al., 2015)	27	A: 89% B: 11%	HBV: 56% HCV: 30%	2.5 cm/ inoperable HCC	60/20 66/22 72/24	3Y: 80% 5Y: 64%	3Y: 17% 5Y: 0%	3Y: 56% 5Y: 42%	Gr 2: 7.4%
(Sugahara et al., 2009)	35	A: 80% B: 20%	HBV: 9% HCV: 86%	6 cm/ with PVT	72.6/22	2Y: 91% 5Y: 91%	2Y: 46% 5Y: 20%	2Y: 48% 5Y: 21%	Gr3: 9%
(Lee et al., 2014)	27	A: 67% B: 33%	HBV: 74% HCV: 15%	7 cm/ with PVT	50-66/20-22	1Y: 71% 2Y: 62%	N/A	1Y: 56% 2Y: 33%	11%

CP=Child-Pugh; PBT=proton beam therapy; RBE= relative biologic effectiveness; HBV=viral hepatitis B; HCV=viral hepatitis C; GI=gastrointestinal; HCC=hepatocellular carcinoma; PVT=portal vein thrombosis; N/A=not applicable; Y=year

Meta-analyses and systematic reviews suggest that PBT is the most effective radiotherapy option for primary liver cancers (Durante, 2015, Qi et al., 2015, Granovetter, 2016, Durante et al., 2017). In Japan, excellent outcomes of PBT in treating liver cancer patients were clearly elucidated (Igaki et al., 2018). However, despite apparent dosimetric benefit together with promising clinical results, there are several limitations in the use of PBT in liver cancers including cost-effectiveness and limited facilities of PBT centers. Therefore, appropriate patient selection is of paramount importance.

Selective use of proton beam therapy in liver cancer

A randomized controlled trial (RCT) between PBT *versus* standard XRT is expected as level I evidence of medical practice. But several obstacles in conducting RCT in this situation exist. First, the "equipoise", or balanced uncertainties bring about ethical compromise because inferior outcomes from one arm can be predicted (Hellman and Hellman, 1991). Given that the same tumor dose distribution results in the same tumor control probability, PBT has been shown to be superior to XRT in terms of reducing unnecessary radiation dose to surrounding normal tissues, known as organs-at-risk. In the other words, PBT likely provides the benefit in terms of reduction of normal tissue complication probability (NTCP). As a result, it is predictable that XRT has inferior dose distribution and might lead to poorer clinical outcomes. Second, greater cost of PBT results in limited availability, capacity and accessibility to conduct RCT. For example, Phase IIb RCT of PBT *versus* intensity-modulated radiotherapy (IMRT) for esophageal cancer at MD Anderson Cancer Center encountered this problem as the insurance company denied the reimbursement in 17 out of 51 patients (33%) in those who were randomized into PBT group. It can lead to slow accrual and, even, early termination of the study. Third, some late radiation-induced complications have long latency period, for example cardiovascular complications and radiation-induced secondary malignancy and take many years to obtain the useful information for treatment decision. At the time of mature study, PBT will eventually out-of-date and will not change any practice. To summarize, RCT might not always be the sole basis of making medical decision in the radiation oncology society where evolving technologies are rapidly developed (Goitein and Cox, 2008). Another alternative method in evidence-based medicine is model-based approach which is currently captivating due to its convenience in development and implementation in allocating

proper treatment modality for individual patient (Langendijk et al., 2013, Shirato et al., 2018, Widder et al., 2016).

Model-based approach

With the aims of appropriate patient selection, decisive model with tailor-made criteria and clinical validation has been expected to select patients who potentially benefit from PBT. In order to establish the selective use of PBT in cancer patients, it is necessary to translate a favorable dose profile of PBT into a reduction of clinically significant toxicities. It has been known that reducing dose to organs-at-risk is the key feature of PBT. Accordingly, NTCP model is applied to estimate the probability of developing toxicity and compare risk of radiation-induced complications between radiotherapy modalities, e.g. PBT *versus* conventional XRT.

A group of researchers in University of Groningen, the Netherlands firstly introduced an approach to select patients most likely to benefit from PBT based on NTCP models, referred as NTCP model-based approach. They proposed the two-phase approach consisting of Phase α: model-based indications: step 1) NTCP model development with prospective data collection and multivariable analysis, step 2) *in silico* planning comparative studies, step 3) estimation of clinical benefit and translation of dose reduction into NTCP reduction (ΔNTCP), and Phase β: clinical validation using prospective observational studies (Langendijk et al., 2013). Successfully, the Dutch health council and the Dutch health care insurance board have finally adopted this approach in 2015. According to the Dutch society of radiotherapy and oncology consensus, PBT is eligible for patients with ΔNTCP of ≥10% for grade 2 toxicity, ≥5% for grade 3 toxicity and ≥2% for grade 4 toxicity (Grau, 2013).

NTCP model comparing PBT *versus* XRT is profoundly attractive in head and neck cancers. *In silico* biological effect analysis for intensity-modulated proton therapy (IMPT) *versus* IMRT showed that a subgroup of patients with upper head and neck tumor location may benefit from IMPT, determined by ΔNTCP ≥10% (Jakobi et al., 2015a). Additionally, use of IMPT for the whole treatment course resulted in ΔNTCP ≥10% in 50% of patients compared to IMRT (Jakobi et al., 2015c). Therefore, NTCP model-based approach can facilitate the preselection of subgroups of patients who potentially exhibit benefit from PBT.

The utilization of *in-silico biomarker* in selecting external radiotherapy was also demonstrated in Japan. Tamura et al. applied Schneider's model to determine lifetime attributable risk of developing secondary cancer in pediatric patients. They found that PBT contributed a significant reduction of risk to only one-third compared to IMRT. Accordingly, Japanese national health insurance has approved PBT in pediatric cancer since 2016 (Schneider et al., 2011, Tamura et al., 2017). In liver tumors, Hokkaido University firstly demonstrated a dosimetric benefit of spot-scanning proton therapy over IMRT for HCC with nominal diameter more than 6.3 cm regarding predicted risk of RILD using Lyman NTCP model, 6.2% for PBT *versus* 94.5% for IMRT. (Toramatsu et al., 2013). However, NTCP model-based approach using ΔNTCP to select PBT in liver cancer patients has never been studied elsewhere**.**

There are some limitations of the NTCP model-based approach for the selective use of PBT. Firstly, there were differences between patient population from which model was derived and patient population for which model was applied. Physical and biological characteristics of patients, diseases and treatments were heterogenous among population. For example, a study from MD Anderson Cancer Center studied the validity and robustness of XRT-derived NTCP models in head and neck cancer patients who were treated with PBT. A reduction of XRT-derived NTCP model efficacy in PBT-treated patients was demonstrated (Blanchard et al., 2016). Secondly, preexisting NTCP models were developed using data from a group of patients, about 100-200 patients, which did not represent the whole population. Moreover, there are considerable sources of model uncertainties that affect the accuracy of model-based approach for patient selection for PBT (Bijman and Breedveld, 2017) as well as the ΔNTCP values. Altogether, model validation and performance as well as uncertainties of NTCP and ΔNTCP should be assessed to improve reliability and confidence of model-based patient selection strategy for PBT.

NTCP model for prediction of RILD

The objective of NTCP model is to translate dosimetric input into a clinically relevant outcome. In other words, "irradiated dose to organ" is depicted as "a probability or risk of developing toxicity of that organ". Various mathematical models have been proposed to estimate the risk of complication as a function of assorted parameters of the treatment plan, tissue architecture and radiosensitivity (Marks et al., 2010). In this study, we mainly focused on NTCP models for prediction of RILD in primary liver cancer patients.

(1) Lyman-Kutcher-Burman NTCP model

Lyman model assumed a sigmoid dose-response relationship between uniform dose irradiated to liver and risk of RILD. For clinical implementation, the Kutcher-Burman effective volume (V_{eff}) was integrated to account for non-uniform irradiation, so called Lyman-Kutcher-Burman (LKB) model (Lyman, 1985, Kutcher and Burman, 1989). In the LKB model, a power-law relationship with a volume effect (n) is used for dose-volume histogram (DVH) reduction into a single parameter. The LKB NTCP model is expressed as follows:

$$\text{NTCP} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2}\, dx \qquad (1)$$

where

$$\text{t} = \frac{D - TD_{50}(v)}{m \cdot TD_{50}(v)} \qquad (2)$$

$$TD_{50}(1) = TD_{50}(v) \times v^{-n} \qquad (3)$$

$$v = \sum_i v_i \left(D_i / D_{ref}\right)^{\frac{1}{n}} \qquad (4)$$

where, $TD_{50}(v)$ is the 50% tolerance dose for uniform distribution of partial volume (v).

$TD_{50}(1)$ represents 50% tolerance dose for whole organ.

m is the steepness of dose-response curve at $TD_{50}(1)$.

n represents the volume effect which relates $TD_{50}(1)$ to $TD_{50}(v)$ in power law relationship as shown in Equation (3). For liver, which is considered a parallel organ, n is very close to 1.

v is the V_{eff} and v_i is the irradiated volume with dose D_i in the i^{th} bin.

D_{ref} is the reference dose which is maximal dose in the DVH.

According to the fact that a true dose distribution is non-uniform, D represents the equivalent uniform dose (EUD) which is the dose that, when given uniformly to the entire organ, would produce the same NTCP as the non-uniform dose distribution (Niemierko, 1997, Luxton et al., 2008). The EUD can be calculated by

$$\text{EUD} = \left(\sum_i v_i D_i^{\frac{1}{n}} \right)^n \qquad (5)$$

where, v_i is the irradiated volume with dose D_i in the i^{th} bin and n represents the volume effect.

However, the 3 parameters in the LKB model ($TD_{50}(1), m, n$) were individually determined by institutions where different dose fractionation schemes were applied. Therefore, Tome et al. suggested the method to correct for different fractionation protocol among studies by conversion of normal liver cumulative DVH dose-bins to a fraction-size equivalent dose (FED) (Tome and Fenwick, 2004) which was described as:

$$FED_{\alpha/\beta}^{f_s} = \frac{Nd\left(1+\frac{d}{\alpha/\beta}\right)}{1+\frac{f_s}{\alpha/\beta}} \qquad (6)$$

where, f_s denotes a reference fraction size or the dose per fraction from which the modeled NTCP data was derived.

d is the physical dose per fraction.

N is the total number of fractions.

α/β is the ratio of linear-quadratic model parameters for the organ, ranging from 2-3 Gy for liver. (In this study, we use α/β =2.0 Gy)

The mean FED is considered as EUD which is expressed as:

$$Mean\ FED_{\alpha/\beta}^{f_s} = \sum_{i=1}^{N_b}\left[FED_{\alpha/\beta}^{f_s}\right]_i v_i \qquad (7)$$

where, N_b is the total number of dose-bins in the differential DVH.

$\left[FED_{\alpha/\beta}^{f_s}\right]_i$ is the FED in the ith dose-bin.

v_i is the partial volume associated with the ith dose-bin.

Several studies previously reported the parameters of LKB NTCP models for predicting RILD as shown in Table 3.

Table 3 Previous studies of Lyman-Kutcher-Burman NTCP models for prediction of RILD

Study	No. of pt (No. of RILD)	Subgroup (n)	Parameters (95%CI) or ±SD			
			TD_{50} (1) (Gy)	*m*	*n*	*f*
(Dawson et al., 2002)	203	entire group (203)	43.3 (41.9-52.8)	0.18 (0.14-0.24)	1.1 (0.88-1.6)	-
	(19)	primary liver cancer (84)	39.8 (38.8-41.1)	0.12 (0.07-0.25)	0.97 (0.69-2.3)	-
(Cheng et al., 2004)	89	whole (89)	49.4 (42-60.5)	0.39 (0.28-0.58)	0.35 (0.20-0.58)	-
	(17)	HBV carriers (65)	50 (42-62)	0.4 (0.27-0.73)	0.26 (0.1-0.48)	-
		HBV noncarriers (24)	46.1 (15- >70)	0.31 (0.18-0.60)	0.86 (0.33- >2)	-
		Child Pugh-A (68)	48.9 (42-59)	0.3 (0.22-0.45)	0.35 (0.21-0.58)	-
		Child Pugh-B (21)	38.7 (33.9-45.1)	0.22 (0.10-1.52)	0.23 (0.14-0.36)	-
		HBV+Child Pugh-A (51)	47.4 (42-57.2)	0.3 (0.20-0.52)	0.3 (0.17-0.50)	-
(Xu et al., 2006)	109	Child Pugh-A (93)	40.5	0.28	1.1	
	(17)	Child Pugh-B (16)	23	0.43	0.7	
(Tai et al., 2009)	414	Child Pugh-A (336)	40.3 ±8.4	0.36 ±0.09	1*	0.156±0.074
	(57)	Child Pugh-B (78)	23.9 ±5.3	0.41 ±0.15	1*	0 ±0.04

**n was fixed to be 1.0 for model fitting due to the limited number of data points.*

RILD=Radiation-induced liver disease; HBV=hepatitis B virus; CI=confidence interval; SD=standard deviation; TD_{50}(1)=whole liver tolerance dose associated with 50% probability of RILD; m=steepness of the dose-response at TD_{50}(1); n=volume effect; f=fitting parameter

(2) Multivariable NTCP model

In routine clinical practice, a single DVH cut-point is accepted to evaluate the integrity of treatment plan. For instance, according to Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) recommendation, volume of normal lung tissue receiving dose of ≤20 Gy (V_{20}) is suggested to be ≤30% to avoid radiation-induced pneumonitis and mean dose to normal liver ≤30-32 Gy is recommended to avoid RILD, and so on. Not only dosimetric parameters are predictive markers for treatment-related toxicities, but clinical and biological factors also substantially affect the risk of complication. For example, liver cirrhosis and chronic viral hepatitis significantly increased risk of developing RILD (Cheng et al., 2004). In 2006, El Naqa et al. initially proposed the application of data mining and model building approach to predict treatment outcomes by mixing clinical and dose-volume factors using a logistic regression framework, so called multivariable modeling (El Naqa et al., 2006). They concluded that multivariable models that do not overfit were more predictive than single-variable models.

The logistic regression multivariable NTCP model can be expressed as:

$$NTCP = \frac{1}{1+ e^{-S(x)}} \qquad (8)$$

$$S(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad (9)$$

where, β_0 is a constant value.

β_i ($1 \le i \le n$) is regression coefficient for the i^{th} covariate (x_i)

Multivariable NTCP model has currently been widely adopted in head and neck cancers (Beetz et al., 2012a) and conjointly applied with model-based approach for selecting patients for PBT (Jakobi et al., 2015a, Langendijk et al., 2013). However, these has never been multivariable NTCP models for predicting RILD in liver cancer.

Purpose of this study

The purpose of our study is to steer towards the selective use of PBT using NTCP model-based approach considering NTCP and ΔNTCP uncertainties for primary liver cancer patients.

Chapter 1 describes the methodology to obtain the estimates of NTCP and ΔNTCP uncertainties based on LKB NTCP model by delta method.

Chapter 2 described the estimates of NTCP and ΔNTCP uncertainties based on LKB NTCP model in subgroups of patients stratified by biological susceptible characteristics.

Chapter 3 describes the novel NTCP model for RILD based on multivariable logistic regression analysis.

Chapter 4 summarizes the findings from this study and demonstrates possible future works followed by problems to be solved.

Chapter 1

Assessing uncertainty of normal tissue complication probability difference: radiation-induced liver disease in liver tumor patients treated with proton vs x-ray therapy

Introduction

PBT has emerged as a promising radiotherapy modality due to its favorable physical properties allowing dose escalation to the tumor with minimal unnecessary dose to normal tissue. With the aims of appropriate patient selection for PBT, a group of researchers at University of Groningen, the Netherlands, firstly introduced the model-based approach to define patients who potentially benefit from PBT compared to XRT by using NTCP model to estimate the risk of developing toxicity (Langendijk et al., 2013). Langendijk et al. and Jokabi et al. investigated the ΔNTCP obtained from *in silico* planning comparing IMPT *versus* IMRT in head and neck cancers (Jakobi et al., 2015a, Langendijk et al., 2013). The selective use of PBT relies on a predefined ΔNTCP threshold, for example 10% for grade 2 toxicity and 15% for total clinical benefit according to the Dutch society of radiation oncology guidelines (Cheng et al., 2016). However, these models were derived from the statistical assumption of a small subset of patient population resulting in model uncertainty which affects the accuracy for PBT selection (Bijman and Breedveld, 2017). In other words, underestimation of ΔNTCP can lead to the loss of opportunity to benefit from PBT while overly cautious practice might cause unnecessary use of this high cost treatment. Therefore, uncertainty around NTCP and ΔNTCP values is necessary for better clinical implementation in a general population.

To develop a reliable and generalized NTCP model, large patient cohorts from multi-institutional dataset further enhance the model accuracy. However, the access to individual DVH data might be limited due to ethical or technical issues in local institutions. Recently, Wedenberg reported the use of dose-bin data obtained from literatures to assess the uncertainty in the estimated dose-response relation for radiation myelopathy and pneumonitis using statistical bootstrap analysis (Wedenberg, 2013). The advantage of dose-bin method is its practically convenience because the data needed for analysis are the number of patients and 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 to individual data for large-

volume data sharing among institutions. However, the accuracy of using this type of data has yet to be warranted.

Chapter 1 described the methodologies of assessing ΔNTCP with uncertainty, 95% confidence interval (95%CI), for comparison of the estimated risk of RILD occurrences between XRT and PBT. Additionally, we compared the results of various methods/algorithms which were used to derive ΔNTCP with uncertainty from individual patient dataset as well as dose-bin dataset.

Methods

NTCP model for RILD

RILD is a dose-limiting toxicity of liver radiotherapy. The most commonly used NTCP model for RILD is LKB NTCP model (Lyman, 1985, Kutcher et al., 1991) with 3 parameters, $TD_{50}(1)$, m and n, as described in equation (1) – (4) above. To account for non-uniform irradiation to the organ, the EUD was applied as in equation (5) (Luxton et al., 2008, Niemierko, 1997). For a parallel organ like liver, volume effect was large (n close to 1) so, in this study, mean liver dose (MLD) referred to EUD calculated from n=1.0 in the LKB model.

Virtual patient dataset generation

Three types of dataset were generated, i.e., reference dataset, dose-bin dataset and bootstrap dataset. They were utilized to illustrate different algorithms to identify NTCP and ΔNTCP with uncertainties. In this study, we created a virtual patient dataset mimicking the Michigan data (Dawson et al., 2002, Randall K. Ten Haken) in Monte Carlo fashion.

(1) Reference dataset

First, MLDs were assigned by random generation algorithm so that its statistics (minimum, maximum, and median MLD) were similar to the statistics of the 203 patients of the Michigan data. In this step, MLD was regarded as a proxy of EUD which were calculated with n=1.1 according to the Michigan's NTCP curve. Then, the probability (NTCP value) was calculated using the Michigan's NTCP parameters, $TD_{50}(1)$=43.3 Gy and m= 0.18 for each of 203 virtual patients. Subsequently, occurrence (adverse event=1) or non-occurrence (adverse event=0) was assigned by Monte-Carlo fashion; i.e., adverse event=1 when a random number (uniformly generated within the range of 0 to 1) exceeded the NTCP at "his/her" EUD, and vice versa. After that, for some sets

of 203 virtual patient data generated from the difference random seeds, again statistics were checked if it was similar to the Michigan data. Finally, a set of data that was nearly identical to the Michigan data were selected as a virtual reference dataset as shown in Figure 2.

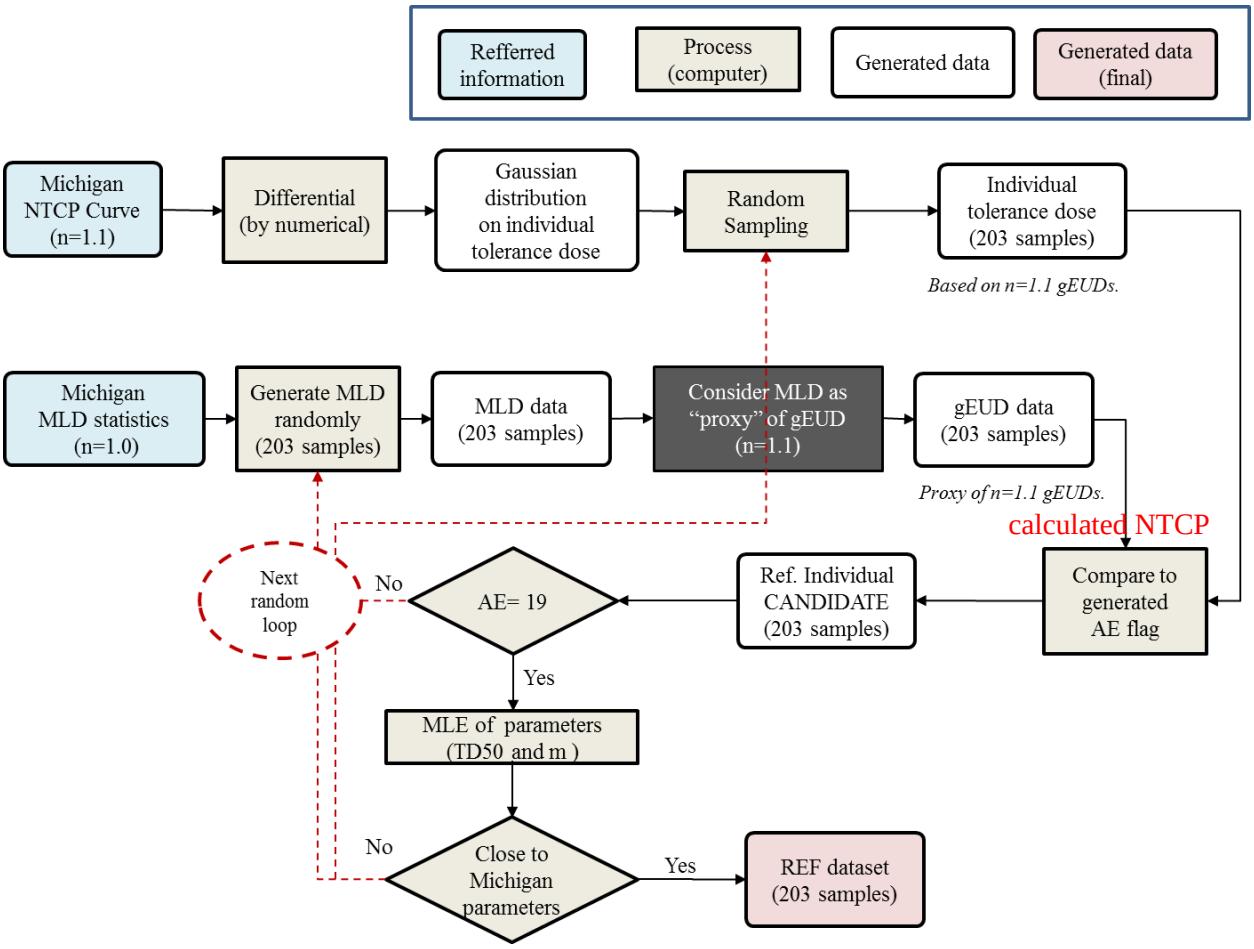


Figure 2 Virtual patient generation for reference dataset

Abbreviation: NTCP=normal tissue complication probability; MLD=mean liver dose; gEUD= generalized equivalent uniform dose; AE=number of adverse events; MLE=maximal likelihood estimation; $TD_{50}(1)$= the 50% tolerance dose for whole organ; n=volume effect; REF dataset =reference dataset

(2) Dose-bin dataset

Reference dataset was then organized by dividing the dose level into equal interval of 5 Gy and counting the number of RILD cases (adverse event) and total number of patients in each dose level/bin. An example of dose-bin dataset was given in Table 4.

Table 4 Organized reference dataset into dose-bin dataset

Dose-bin (Gy)			Adverse event	No. of Patients
Bin	Lower	Upper		
1	15.0	20.0	0	15
2	20.0	25.0	0	32
3	25.0	30.0	1	56
4	30.0	35.0	4	53
5	35.0	40.0	8	32
6	40.0	45.0	6	15

(3) Bootstrap dataset

To check the stability of results and estimates of CI, bootstrapping performs inference about population data from sample data. A newly synthesized individual dataset was generated from dose-bin dataset using bootstrapping technique. In this case, a thousand of bootstrap replicates (a thousand of individual patient dataset) were produced 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 (203 patients) from the same population, referred as bootstrap dataset. A 1,000 bootstrap dataset contributed to 1,000 different values of $TD_{50}(1)$ and m as well as their variability (variance and covariance).

Determination of LKB NTCP model parameters

Patient dataset (reference, dose-bin, bootstrap dataset) comprised the input data to determine LKB NTCP model parameters using maximum likelihood estimation where $TD_{50}(1)$ and m were adjusted to maximize the probability of predicting adverse event for those who experienced RILD and maximize the probability of predicting no adverse event for those who did

not (Martin et al., 2014). Subsequently, variance (σ^2) and covariance of $TD_{50}(1)$ and m were obtained from observed Fisher Information Matrix.The approximate 95%CI was evaluated as 1.96 standard deviations of the mean (σ or $\sqrt{variance}$),1.96 x $\sqrt{variance}$, according to the central limit theorem.

Determination of ΔNTCP

With EUD from treatment plan and the estimated LKB NTCP parameters,$TD_{50}(1)$and m, from maximum likelihood estimation, ΔNTCP between XRT and PBT is given by a function:

$$f(\text{XRT,PBT}) = \text{NTCP}_{\text{XRT}}(\text{EUD}_{\text{XRT}}|\text{TD}_{50}(1),\text{m}) - \text{NTCP}_{\text{PBT}}(\text{EUD}_{\text{PBT}}|\text{TD}_{50}(1),\text{m}) \quad (10)$$

where f is a function of ΔNTCP between XRT and PBT

EUD_{XRT} and EUD_{PBT} denotes normal liver dose for a certain patient for XRT and PBT, respectively.

$TD_{50}(1)$is the 50% tolerance dose for uniform distribution whole organ

m is the steepness of dose-response curve at $TD_{50}(1)$

Determination ΔNTCP with uncertainty

Given the function of ΔNTCP and estimated variance and covariance matrix of $TD_{50}(1)$ and m, the delta method was applied (Armitage et al., 2001). Briefly, the delta method is a standard statistical method to obtain the approximation of variance of a function.

$$var(f) = (\partial f/\partial TD_{50})^2\, var(TD_{50}) + 2(\partial f/\partial TD_{50})\,(\partial f/\partial m)\, cov(TD_{50},m) + (\partial f/\partial m)^2\, var(m) \quad (11)$$

where $\partial f/\partial TD_{50}$ is a partial differential of f with central estimates of TD_{50}

$\partial f/\partial m$ is a partial differential of f with central estimates of m

$var(TD_{50})$ is a variance of TD_{50}

$var(m)$ is a variance of m

$cov(TD_{50},\text{m})$ is a covariance between TD_{50} and m

In present study, 4 algorithms for assessing of ΔNTCP with uncertainty (95%CI) were proposed as followed:

Algorithm #1: Determination of ΔNTCP with 95%CI using delta method in the reference dataset ($\Delta NTCP_{REF}$)

Algorithm #2: Determination of ΔNTCP with 95%CI using delta method in the dose-bin dataset ($\Delta NTCP_{DB}$)

Algorithm #3: Determination of ΔNTCP with 95%CI using bootstrap dataset by analysis of the family of parameters and variability with a method introduced by Efron (Efron and Tibshirani, 1993) to identify the mean of $TD_{50}(1), m$, variance and covariance. Subsequently, ΔNTCP with uncertainty was evaluated using the delta method ($\Delta NTCP_{BS1}$).

Algorithm #4: Determination of ΔNTCP with 95%CI using bootstrap dataset by analysis of the family of NTCP and ΔNTCP values ($\Delta NTCP_{BS2}$) (Wedenberg, 2013)

A study scheme is illustrated in Figure 3.

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

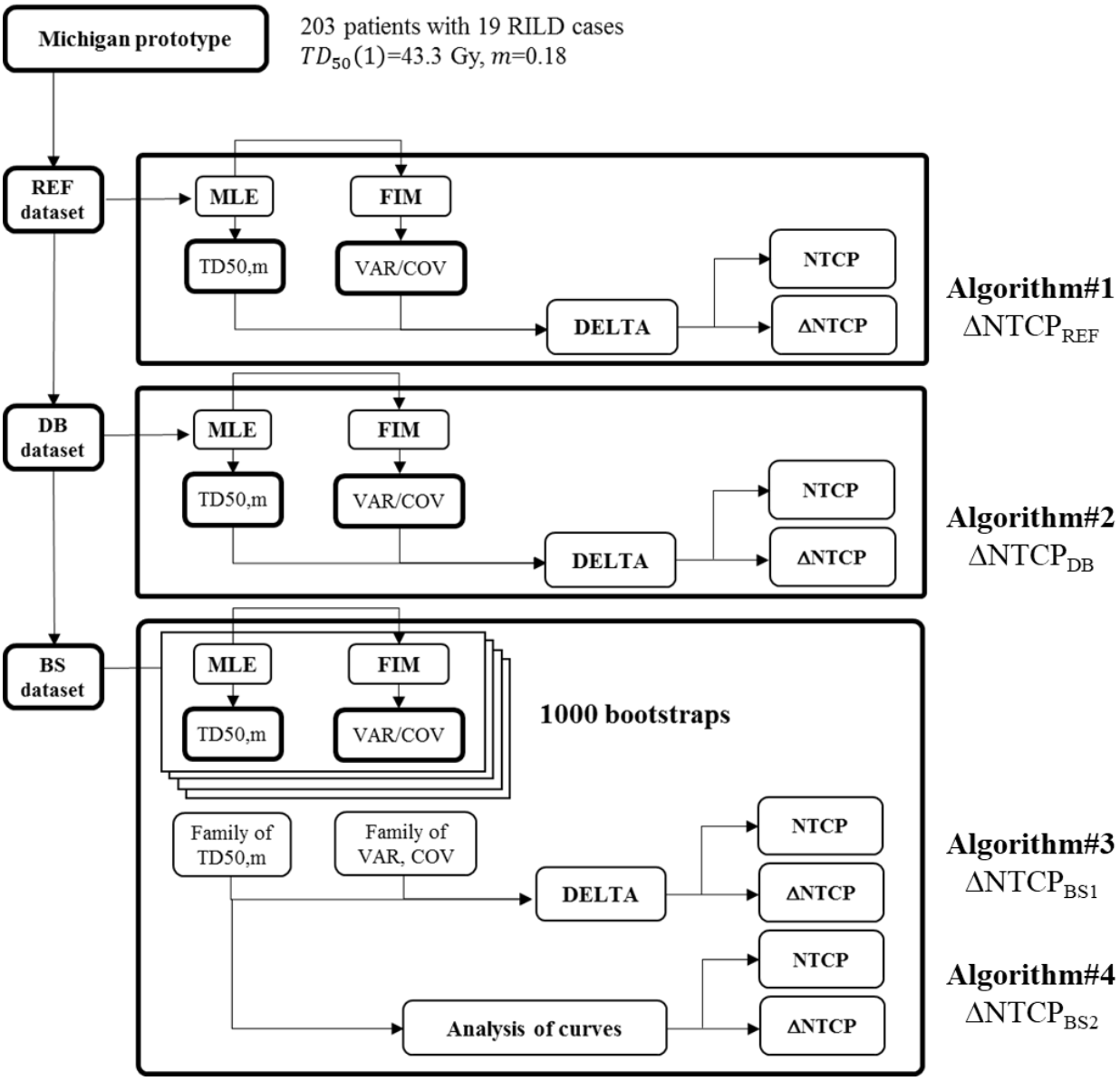


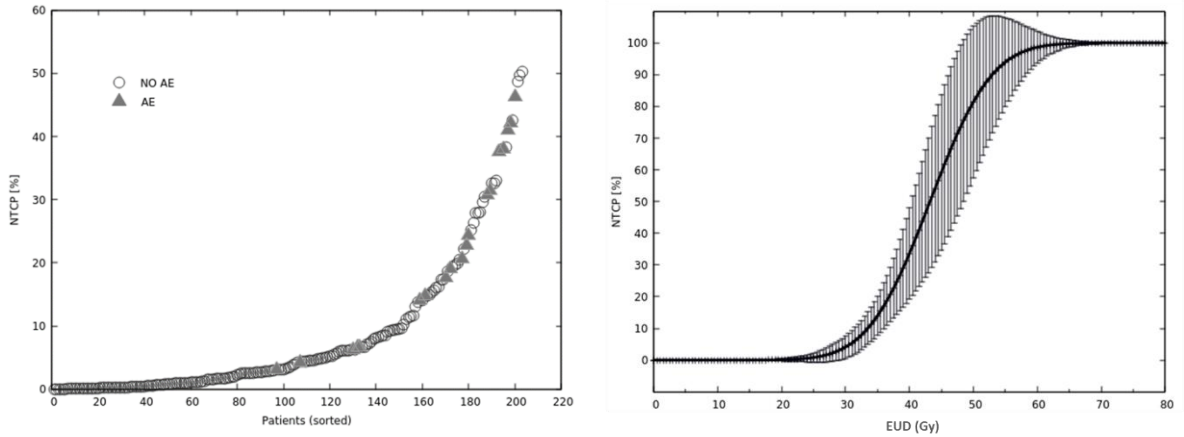
Figure 3 A study scheme illustrated the proposed algorithms to identify NTCP and its difference (ΔNTCP) with 95% confidence intervals

Abbreviation: 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

Results

Estimated LKB NTCP parameters

The virtual reference dataset generated in our study had very similar statistics compared with Michigan study. Average MLD among 203 patients was 29.9 Gy (range, 15.2-43.7 Gy) compared with 32 Gy (range, 14.9-44 Gy) in Michigan's original data. The average MLD was 40 Gy and 28.9 Gy for RILD and non-RILD patients in reference dataset *versus* 37 Gy and 31.3 Gy in Michigan's 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 similar to the original parameters, 43.3 Gy (95%CI 41.9-52.8) and 0.18 (95%CI 0.14-0.24) (Dawson et al., 2002). Figure 4 illustrates the estimated NTCP and EUD of 203 patients with 19 RILD in the reference dataset resembling Michigan's data (Figure 4a) and the estimated NTCP curve with 95%CI as a function of EUD of normal liver considering covariance of (*TD50, m*) (Figure 4b).



(a) (b)

Figure 4 Generated virtual individual patient dataset (reference dataset) of 203 patients with 19 RILD mimicking Michigan's dataset (Fig 4a) contributed to the estimated NTCP curve with 95%CI considering covariance (Fig 4b).

Abbreviation: NTCP=normal tissue complication probability; EUD=equivalent uniform dose; AE=adverse event

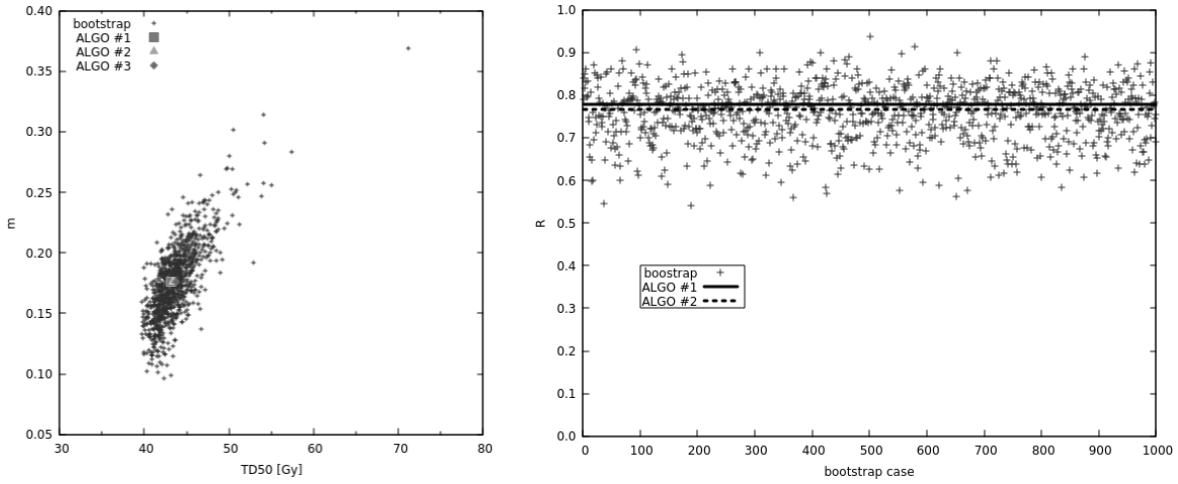
The re-estimated parameters from dose-bin dataset and bootstrap dataset were compared with those from Michigan's and reference dataset as shown in Table 5. Using reference dataset (Algorithm#1), variance of $TD_{50}(1)$ and m were 4.41 and 0.0010, respectively, which were similar to 4.323 and 0.0010 derived from dose-bin dataset (Algorithm #2). The covariance of (*TD50*, *m*) was identified as 0.0519 and 0.502 from Algorithm #1 and #2, respectively. Correlation coefficient calculated from var and cov showed a positively strong relationship between $TD_{50}(1)$ and m with the value of 0.778 in reference dataset and 0.768 and dose-bin dataset. With bootstrapping technique, the distribution of model parameters from 1,000 bootstrap replicates is shown in Figure 5 with the correlation coefficient of 0.8. Due to the nature of uncertainty estimation by curve analysis, variance and covariance of $TD_{50}(1)$ and m were not assessed in Algorithm #4.

Table 5 LKB NTCP parameters with 95%CI and their variance and covariance according to 4 proposed algorithms

Algorithms		**#1**	**#2**	**#3**	**#4***
Dataset	Michigan's data	reference dataset	dose-bin dataset	bootstrap dataset	bootstrap dataset
LKB NTCP parameters					
$\boldsymbol{TD_{50}}$ **(1) (95%CI)**	43.3 Gy (42.9-52.8)	43.2 Gy (39.1-47.3)	43.4 Gy (39.4-47.5)	43.7 Gy (38.7-48.7)	NA
$\boldsymbol{m}$ **(95%CI)**	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					
$\boldsymbol{TD_{50}}$ **(1)**	NA	4.41	4.32	6.47	NA
$\boldsymbol{m}$	NA	0.0010	0.0010	0.0011	NA
Covariance	NA	0.0519	0.0502	0.0606	NA

***** Due to the nature of uncertainty estimation by curve analysis in Algorithm #4, there was no identified representative NTCP curves, $TD_{50}(1)$, m including their variability.

Abbreviation: LKB=Lyman-Kutcher-Burman; NTCP=normal tissue complication probability; CI=confidence interval; $TD_{50}(1)$= the 50% tolerance dose for whole organ; m= steepness of dose-response curve at $TD_{50}(1)$; NA=not available



(a) (b)

Figure 5 Distribution of model parameters from Algorithm #1 (■), #2 (▲) and 1,000 bootstraps (+) with the mean value (♦) (Fig 5a). The correlation coefficients (R) between $TD_{50}(1)$ and m from Algorithm #1 (solid line), #2 (dashed line) and 1,000 bootstrap cases (+) suggested strong relationship between $TD_{50}(1)$ and m (Fig 5b).

Abbreviation: ALGO=algorithm; $TD_{50}(1)$=50% tolerance dose for whole organ; m= steepness of dose-response curve at $TD_{50}(1)$; R=Pearson's correlation coefficient

Comparison of the NTCP curves of Algorithm#1, #2 and #3 showed nearly identical results as demonstrated in Figure 6. Noting that, no representative NTCP curves, $TD_{50}(1)$ and m were identified for Algorithm#4 due to the nature of uncertainty estimation by curve analysis as mention above.

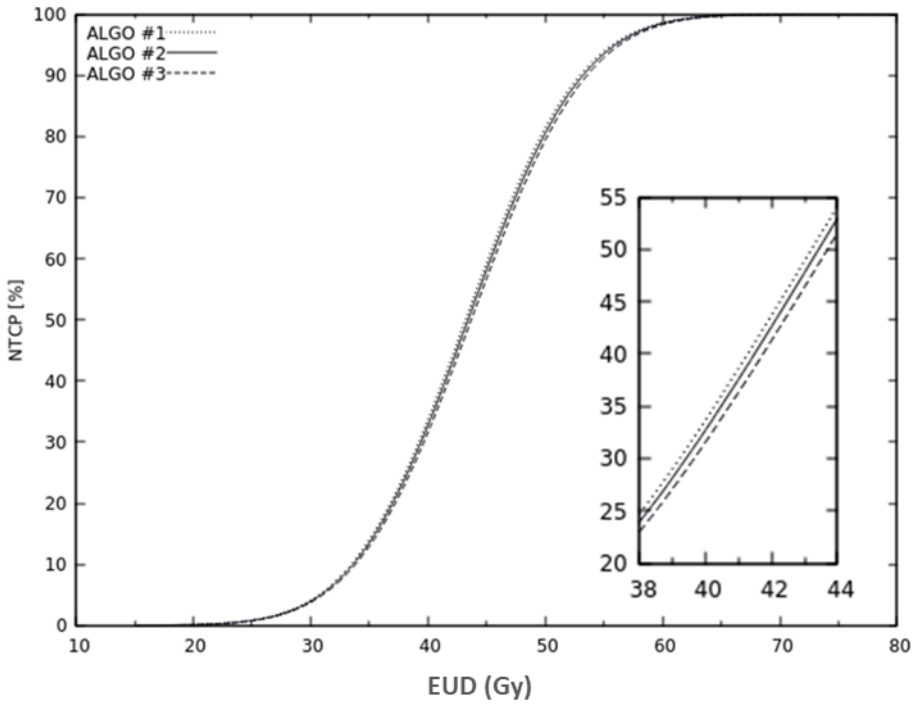
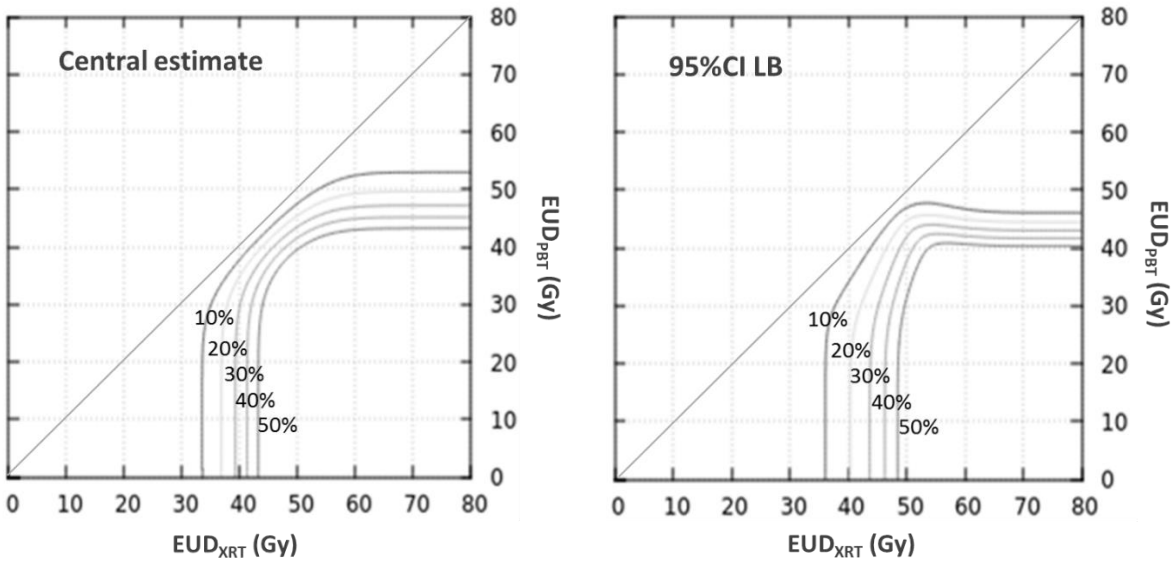


Figure 6 Comparison of NTCP curves from Algorithm#1, #2 and #3

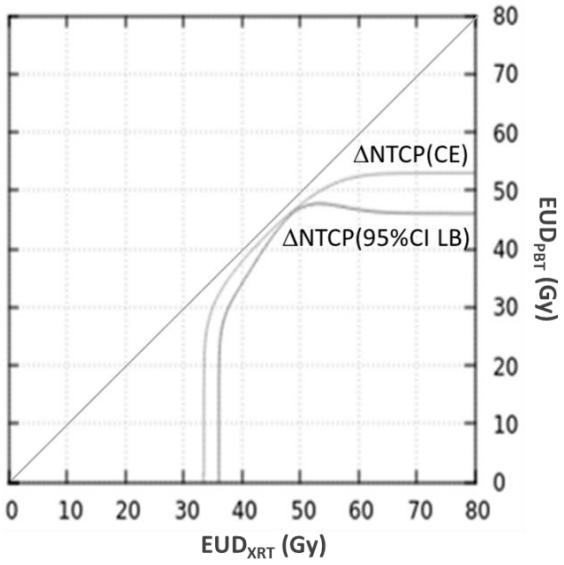
Estimated ΔNTCP with uncertainty

$\Delta NTCP_{REF}$ as a function of EUD between XRT and PBT was shown in Figure 7. Each line represented the iso-ΔNTCP ranging from 10% to 50%. Figure 7a and 7b were contoured for the central estimate and 95%CI lower boundary of iso-ΔNTCP, respectively. With the contours of ΔNTCP considering uncertainty, we can select the patient for PBT based on confidence level. For example, in Figure 7c, using the contour of 95%CI lower boundary, PBT is favored if EUD ≥36 Gy with XRT plan with 95% confidence that ΔNTCP is ≥10%. Contrastingly, using the contours of central estimate, PBT likely provides a potential benefit for a patient who has EUD ≥ 33 Gy with XRT plan with less confidence. In other words, the use of 95%CI lower boundary contour selects less patients for PBT, hence, it is more conservative compared to the central estimate contour. However, it is also suggested that, in the area of 33-36 Gy, the selection of treatment modality should be determined by considering trade-offs between clinical benefits and socioeconomic aspects.



(a)

(b)



(c)

Figure 7 Iso-ΔNTCP contours as a function of EUD between x-ray therapy (EUD_{XRT}) and proton beam therapy (EUD_{PBT}), using delta method in Algorithm #1. Each line represents iso-ΔNTCP with the central estimate (CE) (Fig 7a) and the 95%CI lower boundary (95%CI LB) (Fig 7b), from 10% to 50%. The ΔNTCP-CE and ΔNTCP-95%CI LB at 10% are contoured (Fig 7c).

Abbreviation: EUD= 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

The contours of 95%CI lower boundary of ΔNTCP at 10% among 4 proposed algorithms are compared in Figure 8. When compared to $\Delta NTCP_{REF}$ (Algorithm #1), the contour of $\Delta NTCP_{DB}$ (Algorithm #2) was nearly indistinguishable, so were $\Delta NTCP_{BS1}$ (Algorithm #3) and $\Delta NTCP_{BS2}$ (Algorithm #4). This was true despite varying the percentage of ΔNTCP. However, the contours from bootstrapping techniques yielded a little difference. This might be due to larger covariance found in bootstrap dataset.

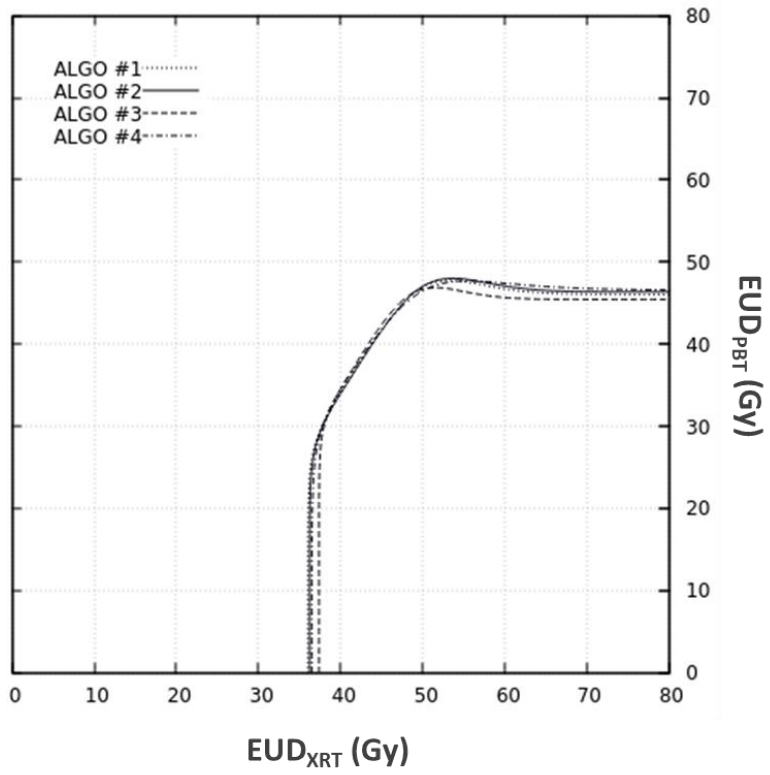


Figure 8 The contours of 95%CI lower boundary of 10% ΔNTCP between XRT and PBT derived from 4 proposed algorithms

Abbreviation: EUD= equivalent uniform dose; XRT=X-ray therapy; PBT=proton beam therapy; ALGO=algorithm

Discussion

Comparing to Michigan's parameters and number of patients reported in the literature (Dawson et al., 2002), the reference dataset was a presumably good representative of the Michigan data despite a small difference in MLD. The re-identified LKB NTCP model parameters were remarkably comparable to the original ones. But variant 95%CI values were observed using different methods to obtain the uncertainty, i.e., profile likelihood method in Michigan's study *versus* variance-based method based on central limit theorem in our study.

With virtual simulated EUD data of 203 patients, the uncertainty of ΔNTCP could be determined by various methods. The 95%CI lower boundary of ΔNTCP provided more conservative threshold and increased confidence for patient selection for PBT compared to the central estimate. Our results also demonstrated that LKB NTCP model parameters and variability derived from dose-bin dataset were very similar to those from individual patient dataset. Additionally, the contour of $\Delta NTCP_{DB}$ was almost identical to the $\Delta NTCP_{REF}$.

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 model-based approach on patient selection for PBT is largely affected by the uncertainty in NTCP models (Bijman and Breedveld, 2017). On the other hand, our study showed that the uncertainty of LKB NTCP model for RILD relies on model parameters variability in terms of variance and covariance. Due to the high correlation between *TD50* and *m*, we firstly assumed that the uncertainty of NTCP curves was better estimated when covariance of *TD50* and *m* was considered.

Mean value of NTCP was generally applied to obtain ΔNTCP (Langendijk et al., 2013, Jakobi et al., 2015a, Cheng et al., 2016). In the present study, we showed that lower boundary of 95%CI considering variance and covariance was more conservative threshold of ΔNTCP and increased confidence in decision-making of the PBT use. In the clinical practice, for a certain patient, comparison of treatment plan between XRT and PBT is usually performed based on dosimetric difference of normal liver DVHs. Using our iso-ΔNTCP curve (Figure 7), if the EUD falls within the area where ΔNTCP is more than the predefined ΔNTCP threshold (the area right to and under the contour), PBT would expectedly be chosen. This 95%CI lower boundary method conservatively selected the patient who potentially benefits from PBT thus the overuse of PBT can be prevented. However, the predefined threshold should be further determined based on clinical outcomes and cost-effectiveness study.

Toramatsu et al. performed a dosimetric comparison between PBT and IMRT in 10 HCC patients with 13 tumors (Toramatsu et al., 2013). We obtained the FED from XRT and PBT plan from Toramatsu's study and applied them into our iso-ΔNTCP contours. We found that the contour of iso-ΔNTCP 95%CI lower boundary at 10% confirmed that tumors with >6 cm would benefit from PBT corresponding with Toramatsu's conclusion as depicted in Figure 9. Notwithstanding, the parameter n was slightly different in our Michigan-resembling NTCP curve (n =1.1) and Toramatsu's (n =0.97).

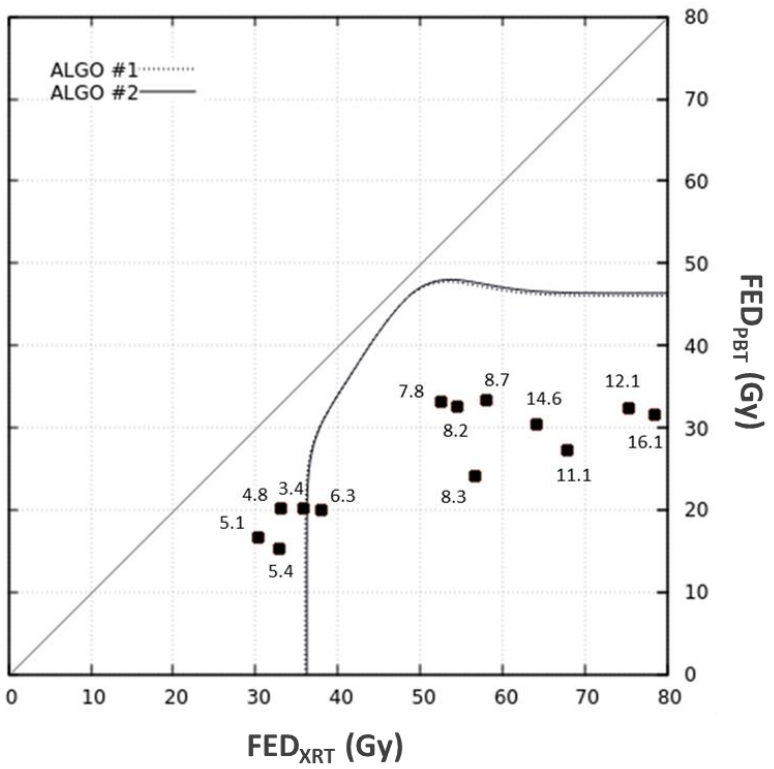


Figure 9 Fraction size equivalent doses of 13 liver tumors in Toramatsu's data (■) plotted with the contour of iso-ΔNTCP 95%CI lower boundary at 10% derived from Algorithm #1 (dashed line), #2 (solid line). The small numbers above denote tumor size in cm.

Abbreviation: FED=fraction-size equivalent dose; XRT=X-ray therapy; PBT=proton beam therapy; ALGO=algorithm

Based on types of data, our study suggests that dose-bin data can be utilized and achieved the same results as individual patient data with more convenience in data collection and sharing between institutions. Algorithm #2 constantly provided the similar results to the Algorithm #1 while Algorithm #3 and #4 were affected by statistical instability and consumed a lot of time.

In this study, due to the lack of individual patient data, we created the virtual simulated patient data based on Michigan study where the fractionation scheme was unique (1.5 Gy twice daily) and concurrent chemotherapy was administered. As a result, our iso-ΔNTCP contours should be cautiously interpreted. Furthermore, it was worth noting that the NTCP model derived from Michigan study dedicated for patients with normal liver function while PBT is theoretically useful for those with impaired hepatic function, e.g. CP-B and CP-C classification (Dawson, 2009). Therefore, the NTCP model needs to include these patient subgroups in order to develop accurate and reliable iso-ΔNTCP decision-making map, especially in a prospective

manner (Prayongrat et al., 2018). Eventually, clinical validation study and cost-effectiveness analysis are expected to identify appropriate ΔNTCP threshold to select patients who potentially benefit from PBT.

In conclusion, the methodology presented in this chapter relies on systematic statistical considerations of 95%CI based on delta method considering variance-covariance matrix thus can be applied to other types of NTCP models and tumors. This might ultimately establish a guideline to appropriately select patients for PBT.

Chapter 2

The normal tissue complication probability model-based approach considering uncertainty for the selective use of radiation modality in primary liver cancer patients

Introduction

PBT has demonstrated excellent tumor control and limited severe adverse outcomes in primary liver cancer patients (Durante et al., 2017, Igaki et al., 2018). However, due to the high treatment cost and limited facilities, appropriate patient selection is mandatory. Considering the lack of RCT comparing PBT with standard XRT, an alternative model-based approach is currently appealing in the radiation oncology field for allocating the suitable treatment modality for individual patients (Grau, 2013, Langendijk et al., 2013, Widder et al., 2016). To identify patients who likely benefit from PBT, a NTCP-based model-based approach can be performed. This approach has been recently studied for the selective use of PBT in head and neck cancer patients (Blanchard et al., 2016, Jakobi et al., 2015a, Cheng et al., 2016).

Apropos of NTCP model for predicting RILD in liver cancers, LKB NTCP model has been well-known. Although the model from Michigan University (Dawson et al., 2002) is de facto standard, there were several studies using LKB NTCP model. But model parameters varied among studies as shown in Table 3 (Dawson et al., 2002, Lawrence et al., 1992, Cheng et al., 2005, Cheng et al., 2004, Tai et al., 2009). The inherent impact of intrinsic biological factors such as liver cirrhosis and viral hepatitis infection on the radiosensitivity of liver tissue has been reported (Cheng et al., 2004), suggesting a different dose-response relationship in individual patients. Additionally, several factors were associated with uncertainty of the model, which affects the accuracy of PBT selection (Bijman and Breedveld, 2017). Underestimating these values can eliminate any benefit gained from PBT, whereas an overly cautious practice might lead to unnecessary treatment.

In Chapter 1, we created the virtual simulated patient data based on Michigan study and proposed the use of delta method considering variance and covariance of model parameters to estimate the uncertainty of ΔNTCP. However, patient population in Michigan study were non-Asian and had unknown viral hepatitis infection status which was common in Asian and considered as one of significant risk factors of RILD. Therefore, we aimed to develop NTCP model for predicting RILD in Asian population and to estimate NTCP and ΔNTCP with uncertainties using the methodology described in Chapter 1. The results would improve the

reliability of the model prediction in the general population and improve the clinical implementation of patient selection for PBT.

Methods

Patients and treatment

Patient eligibility criteria included (i) diagnosis of primary liver cancers (HCC and ICC); (ii) Eastern Cooperative Oncology Group (ECOG) 0-2; (iii) completion of an radiotherapy course; (iv) available 3D dosimetric parameters; and (v) available follow-up data for tumor and liver toxicity, with at least 4 months of follow-up for nontoxicity patients. Patients with progressive disease during a 4-month follow-up were excluded from the study. Demographic data, including age, gender, CP classification, viral hepatitis B and C (HBV and HCV) infection status, baseline liver function and follow-up data were collected. A serum virology diagnosis of HBV chronic infection was defined as positivity for HBV surface antigen (>0.2 ng/mL by chemiluminescent microparticle immunoassay, Abbott Diagnostics, Lake Forest, IL, USA), and HCV infection was defined as a positive anti-HCV status.

All patients received conformal external beam therapy techniques including three-dimension conformal radiotherapy (3D-CRT), IMRT, volumetric modulated arc therapy (VMAT), and SBRT. A planning computed tomography (CT) scan was acquired with patient immobilization. Target volumes were contoured using contrast-enhanced CT scans with a margin expanded to account for subclinical disease, set up uncertainty and respiratory motion. An additional internal target volume was considered for the traditional free-breathing technique until the deep expiratory breath hold technique along with Vision RT® (Varian, Palo Alto, CA, USA) was used to track the patients' surfaces during radiation (since January 2015). Radiotherapy was delivered with a 10-MV or 15-MV linear accelerator. Treatment verification was performed at first fraction then weekly using cone-beam CT.

This study was approved by the institutional review board of Faculty of Medicine, Chulalongkorn University, Thailand (IRB no. 602/60) according to Declaration of Helsinki.

Radiation-induced liver toxicity

After the completion of radiotherapy, a routine follow-up was performed every 1-3 months for treatment-related toxicity or progression of disease. The endpoint of this study was grade≥2 RILT (radiation-induced liver toxicity) according to CTCAE version 4.03, defined as an elevation of liver transaminases (AST and ALT) more than 3xULN, an elevation of ALP

more than 2.5xULN, or an elevation of total bilirubin more than 1.5xULN, in the absence of documented progressive disease. The severity or grading of RILT was scored for each test, and the maximum grade represented toxicity grading of the individual patient. For another commonly used criterion, RILD was classified into classic RILD and nonclassic RILD (Pan et al., 2010). Clinical manifestation of classic RILD included anicteric hepatomegaly, ascites, and elevated ALP more than 2xULN. Nonclassic RILD usually described in patients with pretreatment poor liver function (CP-B/C) or chronic viral hepatitis and characterized as AST or ALT more than 5xULN within 3 months after therapy, or liver function deterioration measured by a decline in CP score by 2 or more.

LKB NTCP model for the prediction of RILT

DVHs of the normal liver tissue (total liver volume minus gross tumor volume) of all patients were obtained from the Eclipse planning system version 8.6 (Varian Medical Systems, Palo Alto, CA, USA). Due to a large volume effect of liver tissue, we fixed the parameter n in the LKB NTCP model at 1.0, known as MLD-based NTCP model (Seppenwoolde et al., 2003). To correct for the difference in treatment regimen, the physical dose distributions were converted to FED using the α/β ratio of 2.0 Gy at a reference fraction size of 2 Gy/fraction (Tome and Fenwick, 2004). MLD in this study was a cumulative result of FED in each dose-bin associated with partial volume associated in that dose-bin as described in Equation (6)-(7).

The dose-response relationship between MLD and CTCAE grade≥2 RILT was modeled by a sigmoid-shaped LKB NTCP model, as in Equation (1)-(4). The methodology to obtain model parameters ($TD_{50}(1)$ and m) as well as their uncertainties was previously described in Chapter 1. Briefly, maximal likelihood estimation was utilized to determine NTCP model parameters then variance and covariance of parameters were obtained from Fisher Information Matrix. The approximate CI was evaluated according to the central limit theorem.

The Hosmer-Lemeshow (HL) goodness-of-fit test was used to evaluate the fitted NTCP model. Model performance was demonstrated using area under the receiver operating characteristic curve (AUC).

Determination of ΔNTCP with uncertainty

With the MLD derived from the treatment plan and the estimated NTCP model parameters ($TD_{50}(1)$ and m), ΔNTCP of PBT compared with XRT is given by ΔNTCP function as in Equation (10). The uncertainty of ΔNTCP considering covariance was estimated

using delta method (Algorithm #1) as described in Chapter 1. Briefly, given the ΔNTCP of certain combinations of MLDs and the estimated variance and covariance matrix, the delta method was applied for estimating the CI of ΔNTCP (Armitage P, 2001).

Statistical analysis

Clinical characteristics and dosimetric parameters between patients with or without toxicity were compared using a χ^2 test or t-test. The correlation between CTCAE grade≥2 RILT and RILD was tested by Spearman rank correlation. Univariate and multivariate analysis was performed using the logistic regression model (multivariate selection: enter method with a probability of entry of 0.20). Due to the multicollinearity between dosimetric parameters (MLD and $V_{x\ Gy}$), only one variable was included in the logistic regression model at each time. A p-value of <0.05 was considered statistically significant. OS was estimated from the date of treatment to the date of death or last follow-up using the Kaplan-Meier method.

Statistical analyses were conducted in R version 3.4.2 (Team, 2010) and IBM SPSS statistics (version 22.0, SPSS, Inc., Chicago, IL).

Results

Between June 2007 and February 2017, 239 patients (146 with HCC and 93 with ICC) were eligible for this study. Majority of the patients were male (73.2%) and CP-A (75.3%). HBV/HCV infection was found in 95 patients (39.7%). The median prescription dose was 45 Gy in a daily fraction of 1.8-10 Gy. There were 5 patients treated with whole liver radiotherapy at 8 Gy in a single fraction. Demographic data are shown in Table 6.

Table 6. Patient and treatment characteristics (N=239)

Characteristics	Value (%)
Age	
mean±SD	60.3±12.5 year
Gender	
male	175 (73.2%)
female	64 (26.8%)
Child-Pugh classification	
A	180 (75.3%)
B-C	59 (24.7%)
Viral hepatitis infection	95 (39.7%)
Diagnosis	
hepatocellular carcinoma	146 (61.1%)
cholangiocarcinoma	93 (38.9%)
Radiotherapy technique	
3D-CRT	117 (49%)
IMRT/ VMAT/ SBRT	122 (51%)
Fraction size	
≤3 Gy	180 (75.3%)
>3 Gy	59 (24.7%)
Radiation treatment	
total prescription dose, median (IQR)	45 (30-50) Gy
number of fractions, median (IQR)	10 (10-25) fractions
dose per fraction, median (IQR)	3 (1.8-3.5) Gy
normal liver volume*, mean±SD	1100.7 (890.7-1418.1) cc

3D-CRT=three-dimensional conformal radiotherapy; IMRT=intensity-modulated radiotherapy; VMAT=volumetric arc radiotherapy; SBRT=stereotactic body radiotherapy; SD=standard deviation; IQR=interquartile range

**normal liver volume=total liver volume minus gross liver volume*

Radiation-induced hepatic toxicity

At the median follow-up time of 10.8 months, a total of 131/239 patients (54.8%) developed CTCAE grade≥2 RILT or liver dysfunction equivalent to CTCAE grade≥2 RILT after radiotherapy, namely, 83/180 (46.1%) in the CP-A group and 48/59 (81.4%) in the CP-B/C group (p<0.001). Among 95 patients with viral hepatitis infection, 60 patients (63.2%) experienced CTCAE grade≥2 RILT, which was significantly higher than the 49.3% of patients without infection (p=0.035). Forty-seven patients were diagnosed with RILD (19.7%): 16/180 patients (8.9%) in the CP-A group *versus* 31/59 patients (52.5%) in the CP-B/C group (p<0.001). There was a correlation between CTCAE-RILT and RILD by the Spearman rank test (r=0.393, p<0.001). Patients who developed a CTCAE grade≥2 RILT had shorter median OS which was 8 months *versus* 17.5 months (p<0.001).

In the multivariate analysis, CP classification, normal liver volume, MLD, V_5, V_{10} and V_{20} were significantly associated with CTCAE grade≥2 RILT. CP classification was the strongest predictor with OR of 5.55 (p<0.001). Table 7 summarizes the association between clinical and dosimetric parameters and CTCAE grade≥2 RILT.

Table 7 Univariate and multivariate analysis for clinical and dosimetric variables associated with grade≥2 radiation-induced liver toxicity

Characteristics	RILT		p-value	
	Grade <2 (N=108)	Grade ≥2 (N=131)	Univariate analysis	Multivariate analysis
Age (years)	60.8±12.8	59.9±12.3	0.541	-
Gender			0.542	-
male	77	98		
female	31	33		
Diagnosis			0.017	0.913
HCC	56	88		
ICC	52	43		
Child-Pugh and viral hepatitis subgroups				
CP-A, non-hepatitis	58	44	<0.001	<0.001
CP-A, hepatitis	37	34		
CP-B or C	13	53		
PVT (N=110)	34	76	<0.001	0.034

Characteristics	RILT		p-value	
	Grade <2 (N=108)	Grade ≥2 (N=131)	Univariate analysis	Multivariate analysis
Previous surgery (N=69)	43	26	0.001	0.557
TACE (N=87)	42	45	0.277	-
Chemotherapy (N=40)	18	22	1.000	-
Radiotherapy technique			0.973	-
3D-CRT	53	64		
IMRT/ VMAT/ SBRT	55	67		
Gross tumor volume (cc)	220.6	553.1	0.001	0.171
Normal liver volume* (cc)	1107.6±333.1	1302.7±721.9	0.006	0.209
Mean MLD to liver (Gy)	16.1±7.3	18.0±7.3	0.041	0.017
V_{5Gy} (%)	59.4±23.6	67.8±22.5	0.005	0.006
V_{10Gy} (%)	47.7±22.9	55.8±24.1	0.009	0.025
V_{20Gy} (%)	31.0±18.1	34.9±19.1	0.105	0.149
V_{30Gy} (%)	20.5±14.0	21.9±15.6	0.478	-
V_{40Gy} (%)	12.1±10.5	11.5±11.2	0.675	-

Abbreviation: RILT=radiation-induced liver toxicity; HCC=hepatocellular carcinoma; ICC=intrahepatic cholangiocarcinoma; CP=Child-Pugh classification; PVT=portal vein thrombosis; TACE=transarterial chemoembolization; 3D-CRT=three-dimensional conformal radiotherapy; IMRT=intensity-modulated radiotherapy; VMAT=volumetric arc therapy; SBRT=stereotactic body radiotherapy; MLD=mean liver dose; $V_{x\ Gy}$=volume receiving x Gy

**normal liver volume=total liver volume minus gross liver volume*

NTCP model

For the MLD-based NTCP model, the estimated $TD_{50}(1)$ and m parameters were inconsistent among patient subgroups as summarized in Table 8.

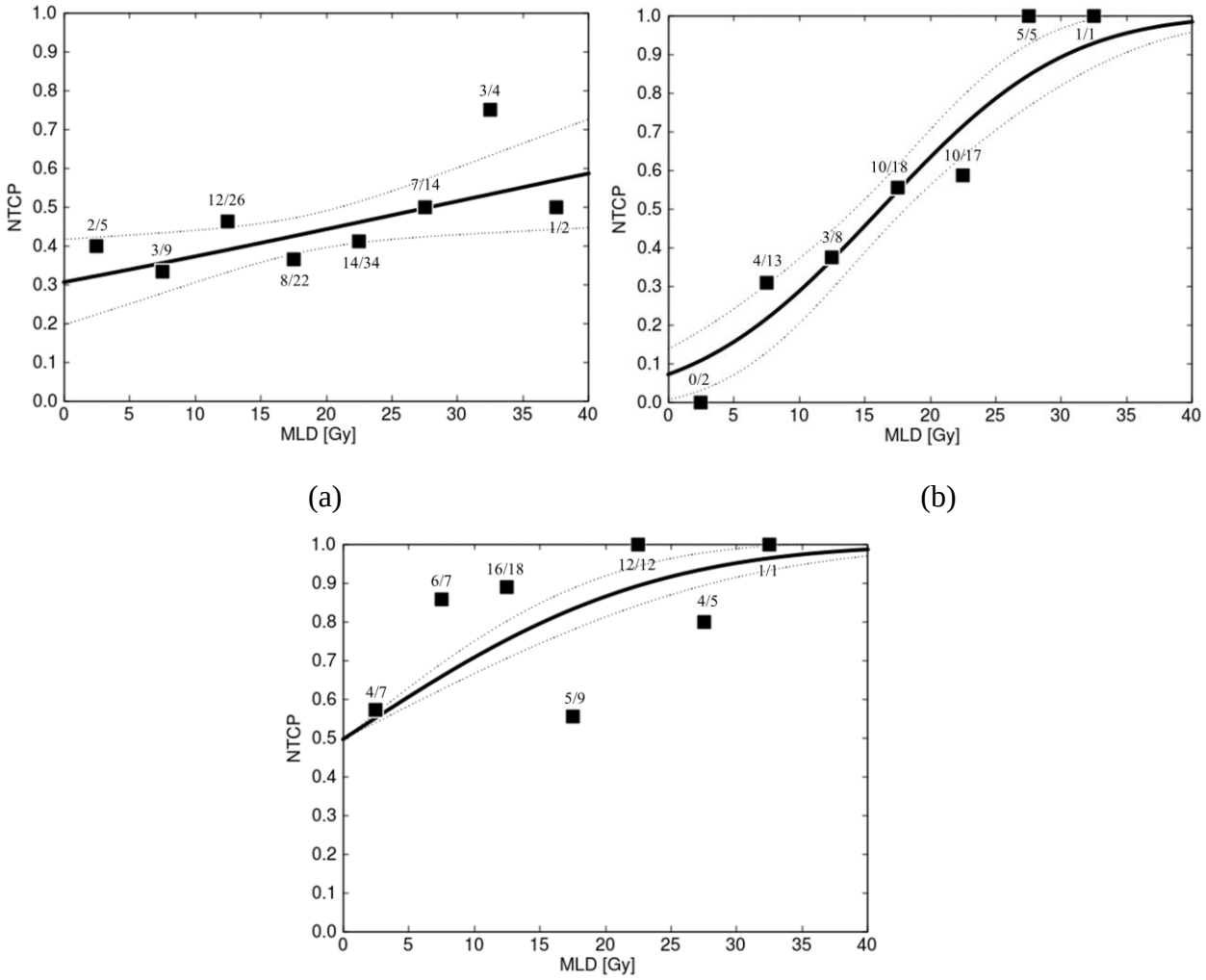
Table 8 Estimated NTCP model parameters for predicting grade≥2 radiation-induced liver toxicity of the whole group and subgroups

	Parameters	
	TD_{50} (Gy)	m
Whole group (N=239)	11.8	3.70
Child-Pugh A, nonhepatitis (N=116)	28.0	1.97
Child-Pugh A, hepatitis (N=64)	16.2	0.68
Child-Pugh B* (N=59)	0.2	99.89

NTCP=normal tissue complication probability; TD50=50% tolerance dose for irradiation to the whole liver; m=steepness of the dose-response curve at TD_{50}

** A TD_{50} slightly closer to zero and a larger value of m can be obtained as another converged maximum likelihood estimation solution, but the shape of the curve and its CI were almost equivalent to this set of parameters.*

CP-A patients were divided into two subgroups: CP-A, the nonhepatitis subgroup (N=116), and CP-A, the hepatitis subgroup (N=64). The analysis was not performed in the CP-B/C subgroup after stratification by hepatitis status due to the limited number of patients. All models in every subgroup fit the clinical data well (χ^2_{HL}=6.90, p=0.547 for the CP-A, nonhepatitis subgroup; χ^2_{HL}=7.50, p=0.483 for the CP-A, hepatitis subgroup; and χ^2_{HL}=10.54, p=0.229 for the CP-B/C subgroup). Overall model performance was acceptable with AUC of 0.714 (95%CI, 0.650-0.779) compared with AUC of 0.569 (95%CI, 0.496-0.642) when using Michigan's parameters (Dawson et al., 2002). Large differences in the NTCP curves were seen among subgroups stratified by CP classification and viral hepatitis infection status as shown in Figure 10.



(a) (b)

(c)

Figure 10 The incidence of CTCAE grade≥2 RILT or liver dysfunction equivalent to CTCAE grade≥2 RILT of Child-Pugh A patients without (Figure 10a) and with (Figure 10b) viral hepatitis infection subgroups and Child-Pugh B/C patients (Figure 10c) subgroups as a function of MLD. The solid line represents the probability of CTCAE grade≥2 RILT according to the NTCP model. The dotted lines represent the 68% CI of the fitted NTCP curve. In each dose-bin (■), small numbers above denote number of patients with CTCAE grade≥2 RILT / total number of patients.

Abbreviation: CTCAE=common toxicity criteria of adverse events; RILT=radiation-induced liver toxicity; *MLD=mean liver dose; NTCP=normal tissue complication probability*; *CI=confidence interval*

ΔNTCP with uncertainty

At a certain combination of MLDs, the mean values of NTCP were adopted to obtain the ΔNTCP. Then, the uncertainty of ΔNTCP was acquired by the delta method, which considered the variance and covariance matrix of the NTCP parameters. Iso-ΔNTCP at 10% is contoured in Figure 11. The area to the right and below this contour line represented the area with ΔNTCP ≥10% which was considered a PBT-benefit area in this study.

To give an example of implementing our ΔNTCP contours in clinical practice, we applied the DVH data of HCC tumors from Toramatsu et al. (Toramatsu et al., 2013). The estimated ΔNTCP with a 68%CI for 5 patients assuming different CP and viral hepatitis status are listed in Table 9. Their tumor sizes according to their FED from PBT and XRT plan are plotted in Figure 11. It was suggested that, assuming a different CP classification and status of viral hepatitis infection, eligibility for PBT were different between subgroups.

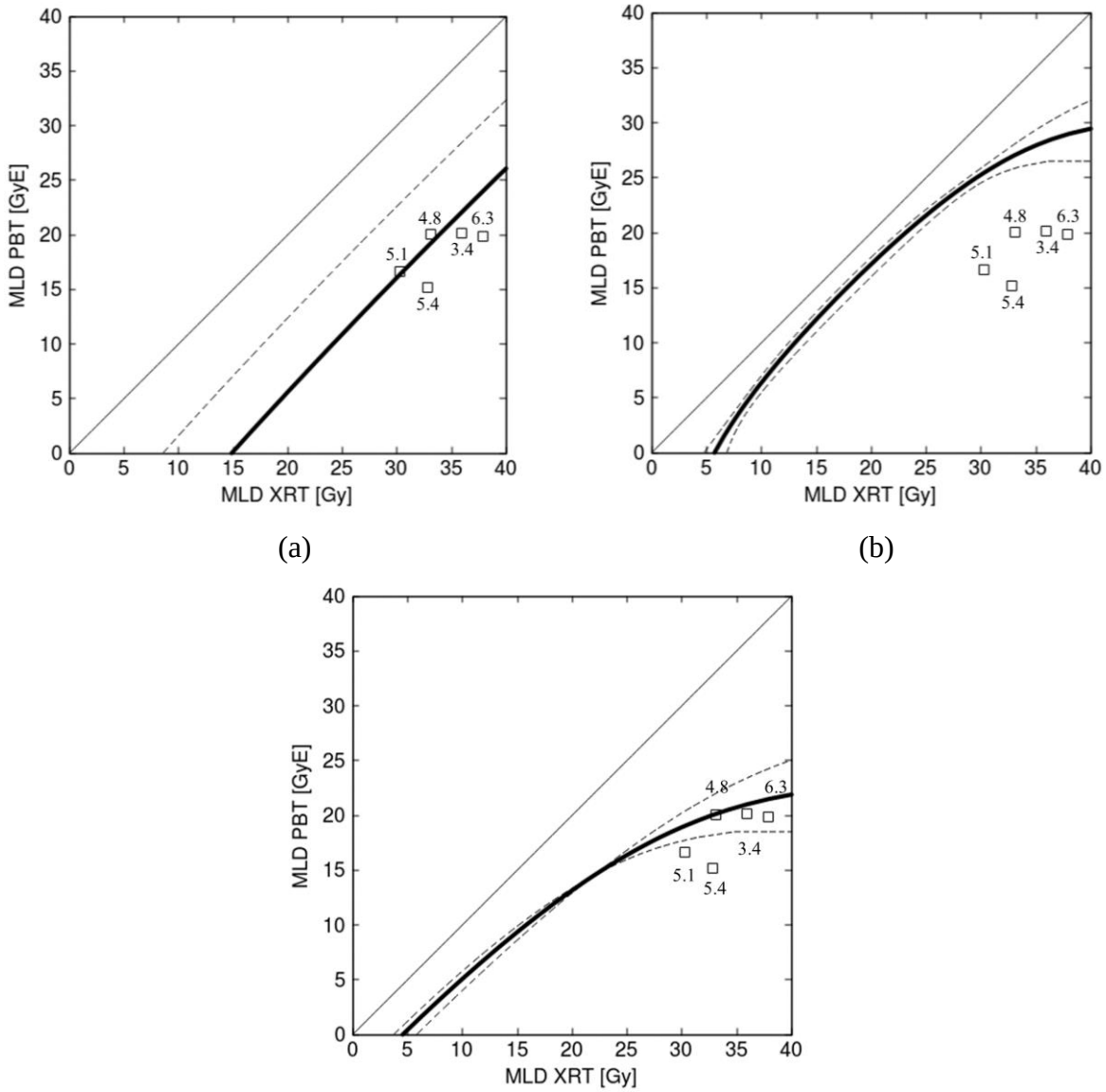


Figure 11 The ΔNTCP of CTCAE grade≥2 RILT between two radiation modalities of certain combinations of the MLD in Child-Pugh A patients without (Figure 11a) and with (Figure 11b) viral hepatitis infection subgroups and Child-Pugh B/C patients (Figure 11c). The diagonal line is the MLD-equivalent line between XRT and PBT. The solid and dotted black lines represent the central estimate and 68% CI of iso-ΔNTCP at 10%, respectively. The small box represents the size of tumors (cm) plotted according to Toramatsu study.

Abbreviation: MLD=mean liver dose size; XRT=X-ray therapy; PBT=proton beam therapy

Table 9 ΔNTCP with a 68% confidence interval for predicting the probability of CTCAE grade≥2 RILT from two treatment modalities assuming a different CP classification and viral hepatitis infection status

Plan ID	Mean MLD IMRT (Gy)	Mean MLD PBT (Gy)	ΔNTCP (68% CI) for CTCAE grade≥2 RILT		
			Child-Pugh A, nonhepatitis	Child-Pugh A, hepatitis	Child-Pugh B/C
1	35.9	20.2	11.5% (1.6%-21.5%)	33% (28%-38%)	11.2% (8.3%-14.1%)
2	33.1	20.1	9.4% (1.2%-17.5%)	30.2% (25.5%-34.9%)	10.1% (7.9%-12.4%)
3	30.3	16.6	10.1% (1.3%-18.8%)	39.1% (31.5%-46.8%)	13.6% (11.9%-15.4%)
4	32.8	15.2	12.9% (1.7%-24.1%)	47.9% (39.2%-56.6%)	17.1% (14.9%-19.3%)
5	37.9	19.9	13% (1.8%-24.1%)	34.2% (28.9%-39.5%)	11.7% (8.5%-15%)

Abbreviation: ΔNTCP=NTCP reduction; CI=confidence interval; CTCAE=common toxicity criteria of adverse events; RILT=radiation-induced liver toxicity; MLD=mean liver dose; IMRT=intensity-modulated radiotherapy; PBT=proton beam therapy

Discussion

Radiation-induced hepatitis is a dose-limiting complication of traditional liver-directed radiotherapy and can lead to deterioration of liver function followed by liver failure and death (Dawson et al., 2002, Pan et al., 2010, Cheng et al., 2002, Tanguturi et al., 2014). Therefore, predicting the risk of toxicity in individual patients with unique clinical characteristics would facilitate treatment decision-making concerning appropriate radiotherapy modality/technique/regimen. Intrinsic biological factors, such as poor hepatic functional reserve and chronic viral hepatitis, were reported to have a potential effect on the susceptibility of liver tissue to irradiation after 3D-CRT (Cheng et al., 2004, Dawson, 2009, Liang et al., 2006). Our study showed the differential influence of biosusceptibility factors to RILT, consistent with these studies. We demonstrated the use of the LKB NTCP model to estimate the risk of CTCAE grade≥2 RILT as a function of MLD and reported new sets of estimated NTCP parameters for CTCAE grade≥2 RILT for specific patient subgroups, namely, the CP-A, nonhepatitis subgroup; the CP-A, hepatitis subgroup; and the CP-B/C subgroup.

To select patients who will gain a clinical benefit from PBT, Langendijk et al. has proposed a stepwise approach, referred to as the model-based approach (Langendijk et al., 2013). This principle has been given attention by the health authorities of the Netherlands (Grau, 2013, Widder et al., 2016) with a threshold for clinical benefit considered as a predicted reduction in the probability of grade≥2 toxicity of >10% (a total reduction of ≥15% for the complication profile) (Cheng et al., 2016). In this study, we contoured the ΔNTCP with a 68% CI at 10% ΔNTCP of developing CTCAE grade≥2 RILT in liver cancer patients as an example of implementing this approach. As shown in Table 9 and Figure 11, assuming a different CP classification and viral hepatitis infection status, the NTCP and ΔNTCP markedly varied and thus may have a potential impact on the decision regarding PBT use. For example, for plan ID 2 (MLD=33.1 Gy for the IMRT plan *versus* 20.1 Gy for the PBT plan), considering a ΔNTCP threshold of 10% reduction of CTCAE grade≥2 RILT, PBT was indicated for patients with CP-A, hepatitis (ΔNTCP=30.2%) and CP-B/C (ΔNTCP=10.1%) but not for the CP-A, nonhepatitis subgroup (ΔNTCP=9.4%). This finding emphasizes the differential benefit from PBT according to patient subgroup and the importance of appropriate patient selection. Of note, Toramatsu et al. applied Michigan NTCP parameters for RILD and corrected fraction size to 1.5 Gy/fraction with an α/β ratio of 2.5 Gy.

Moreover, there are considerable sources of model uncertainties and statistical assumptions from a small subset of patients when developing the model (Bijman and Breedveld, 2017, Schilstra et al., 2001, Wedenberg, 2013). The true population-based dose-response relationship can be given estimated by the upper and lower limits of the CI, which improves the reliability of the model prediction in the general population. In the example in Table 9, if we set the ΔNTCP 68%CI lower boundary at 10% as a threshold, PBT was indicated for all 5 patients in the CP-A with hepatitis subgroup, but none in the CP-A with nonhepatitis subgroup, and two CP-B/C patients (plan ID 3 and 4).

Our study proposed a clinical decision support using NTCP model-based approach to select patients for PBT as presented in Figure 12. Firstly, subgroup-specific NTCP model for predicting RILT was selected according to CP classification and viral hepatitis status. Secondly, planning doses from XRT and PBT were applied in the model to estimate the NTCP values, $NTCP_{XRT}$ and $NTCP_{PBT}$, respectively. Thirdly, the ΔNTCP, defined by $NTCP_{XRT}$ - $NTCP_{PBT}$, was obtained and used to select the appropriate modality according to the predefined threshold (10% in this example) which corresponded with the area to the right and below the contour in Figure 11. It is worth noting that the threshold could be central estimate of ΔNTCP as well as upper or lower boundary of ΔNTCP CI.

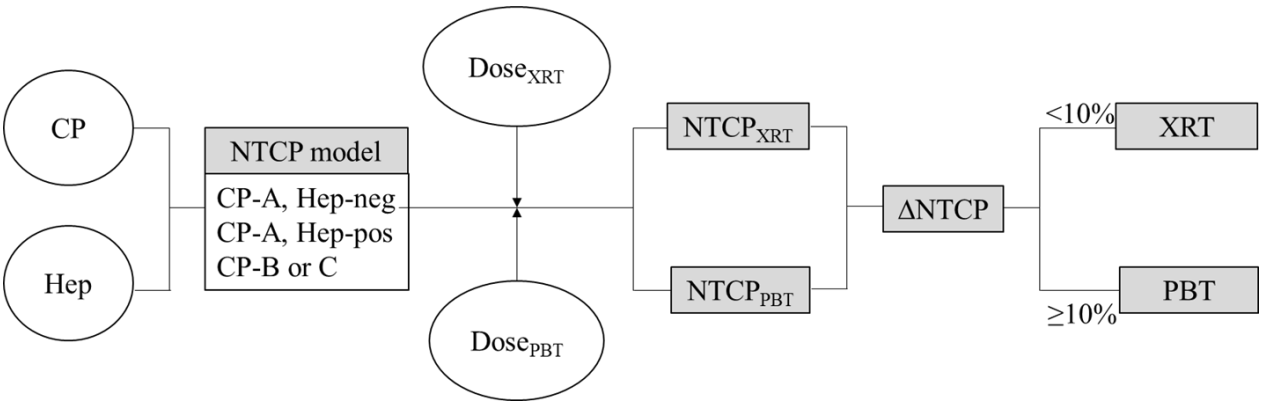


Figure 12 clinical decision support using NTCP model-based approach

Abbreviation: CP=Child-Pugh; Hep=viral hepatitis infection status; XRT=x-ray therapy; PBT=proton beam therapy; NTCP=normal tissue complication probability; ΔNTCP=NTCP reduction

However, appropriate ΔNTCP threshold (10% or other value) as well as proper confidence level (68%CI or 95%CI, upper or lower boundary) for the selective use of PBT is still unknown. Consequently, the desired goal and challenges are to determine the threshold for identifying patients who would most likely benefit from PBT. The physical dose distribution, biological impact of the treatment on tumor and normal tissues, clinical impact on patients' quality of life and socioeconomic aspects need to be considered and optimized (Shirato et al., 2018). As a future research, we should perform cost-effective analysis of our model-based approach to improve quality-adjusted life years of patients and total cost of national medical care.

This study has several strengths. First, the use of the delta method, considering variance and covariance of NTCP parameters, produced ΔNTCP with CI that are less overestimated and more generalizable. Second, unlike previously reported studies on NTCP for RILD (Dawson et al., 2002, Lawrence et al., 1992, Xu et al., 2006, Cheng et al., 2005, Cheng et al., 2004, Tai et al., 2009), our model predicted the probability of CTCAE grade≥2 RILT as a function of the MLD. This outcome is based on a widely accepted standard toxicity criterion that is objective, easily assessable and convenient for data collection. In addition, CTCAE grade≥2 RILT was moderately correlated with RILD but earlier detected, which might rise the concern of treating physician prior to severe changes of liver function. Third, the use of the MLD with the fraction-size normalization method mitigates the limitation of comparing DVH data between various fractionation regimens and eliminates the need for n optimization in the LKB model. However, given the different physical and biological properties between PBT *versus* XRT and the effect of changes in dose distributions on the predictive power of the NTCP model, the XRT-derived NTCP models should be validated prior to the direct comparison of toxicity risks (Blanchard et al., 2016). Thus, multi-institutional data pooling is required, and the use of CTCAE and MLD helps facilitate data sharing among institutions.

The limitations of our study were the lack of data in the high (>45 Gy) dose-bin and the small patient number after stratification into subgroups. It was also noted that some patients had liver dysfunction equivalent to CTCAE grade≥2 RILT despite a very low MLD. This high background, up to 50% in CP-B/C patients, might be due to pre-existing poor hepatic function, reactivation of viral hepatitis or rapid tumor reaction and might not be related to radiation. However, the background was eventually subtracted when calculating the ΔNTCP. In addition,

despite stratification according to two clinical factors (CP classification and viral hepatitis infection status), our NTCP models considered only MLD to predict toxicity outcome. Other possible prognostic or predictive parameters for RILT include tumor size and origin, portal vein thrombosis, and previous treatment (Dawson et al., 2002, Liang et al., 2006) which justify the development of a multivariable model in our future work to accurately predict RILT in various circumstances. Another limitation was a lack of model validation in this analysis. It is important to validate our model using a validation dataset as a next step.

In conclusion, CP classification was the strongest clinical predictive factor for risk of toxicity after liver radiotherapy, followed by viral hepatitis infection. We reported the estimated LKB NTCP model parameters specific to patient subgroups stratified by CP classification and viral hepatitis infection status. NTCP curves with a 68%CI were determined for each subgroup, considering the variance and covariance of parameters. The ΔNTCP at a certain combination of MLDs (including its uncertainty) might be helpful to guide the selective use of PBT.

Chapter 3

Multivariable normal tissue complication probability model for predicting radiation-induced liver disease in primary liver cancer

Introduction

A number of studies have shown the relationship between dose to normal liver and hepatic toxicity based on LKB NTCP model (Dawson et al., 2002, Lawrence et al., 1992, Xu et al., 2006, Cheng et al., 2005, Cheng et al., 2004, Tai et al., 2009) including our study in chapter 2. Besides liver dosimetry, the importance of biosusceptibility factors, such as liver cirrhosis and viral hepatitis infection, on the risk of RILD has also been emphasized (Cheng et al., 2004, Prayongrat et al., 2019). Other possible non-dosimetric predictors included male gender, presence of PVT, prior and concurrent treatments (Cheng et al., 2004, Dawson et al., 2002, Liang et al., 2006, Xu et al., 2006).

NTCP model can potentially guide treatment selection for patients who likely benefit from PBT in terms of reduction in clinically significant toxicities based on predefined ΔNTCP threshold (Langendijk et al., 2013, Widder et al., 2016, Jakobi et al., 2015a). In Chapter 2, LKB NTCP models for predicting liver toxicity were developed for primary liver cancer patients which were stratified into subgroups by CP classification and viral hepatitis infection status. We also revealed the differences of NTCP and ΔNTCP between subgroups . However, apart from CP classification and viral hepatitis infection, there were other potential clinical factors as mentioned above. Therefore, the purpose of Chapter 3 is to develop a multivariable prediction model that combines clinical and dosimetric parameters to identify primary liver cancer patients at risk for developing RILD that can further support decision making to select appropriate treatment.

Materials and methods

Patient population and Treatment

Patient population of this retrospective study composed of 326 primary liver cancer patients treated between December 2006 and September 2018 at the department of radiation oncology, King Chulalongkorn memorial hospital, faculty of medicine, Chulalongkorn University, Thailand. The inclusion criteria and treatment technique were the same as described in Chapter 2.

Briefly, gross tumor volumes were contoured using contrast-enhanced CT scans, and magnetic resonance image (MRI) if available, with a 5- to 10-mm margin expanded to account for subclinical disease, set up uncertainty and respiratory motion. Normal liver was contoured as entire liver subtracted with gross tumor volume. Some patients underwent surgery or TACE prior to radiotherapy. No concurrent chemotherapy was administered.

The study has been approved by local institutional review board (IRB no. 602/60).

Data extraction of clinical and dosimetric parameters

Baseline patient characteristics including age, gender, tumor size, presence of PVT, CP score, HBV/HCV infection status and previous treatments were reviewed. Physical examination, blood tests including liver function test (AST, ALT, ALP, total bilirubin, albumin) and imaging were assessed at baseline and every 1-3 months after completing RT for evaluating treatment-related toxicities and disease progression. Toxicity endpoint was RILD which was classified into **"**classic**"** and **"**non**-**classic**"** RILD as described in detail in previous Chapters**.** In the current study, classic RILD was the endpoint for CP**-**A and negative viral hepatitis infection status patients whereas nonclassic criteria was scored for patients with either CP**-**B/C or positive viral hepatitis infection status**.**

DVHs of all patients were obtained from the Eclipse planning system version 8.6 (Varian Medical Systems, Palo Alto, CA, USA). To simplify the whole DVH into a single measurement and simultaneously account for organ architecture, normal liver DVH was reduced and adjusted for fraction size using Equation (6)-(7) into MLD.

NTCP Modelling and statistical analysis

NTCP model for RILD was developed using data from 280 patients (training dataset) and validated in 46 patients (test dataset) who were independent from the training dataset and treated between January 2015 and September 2018. Clinical and dosimetric parameters between patients with or without RILD were compared using a χ^2test or *t*-test. OS was estimated from the date of radiotherapy start to the date of death or last follow up using the Kaplan-Meier method. A univariate analysis was initially performed for all candidate predictor variables followed by a multivariate logistic regression analysis with bootstrapping technique and forward variable selection. Significant variables were eventually selected in the final model with *p*-value of <0.05

to be considered statistically significant. Due to the multicollinearity between dosimetric parameters, (MLD and $V_{x\,Gy}$), only one variable was included in the logistic regression model at a single time. OR and 95%CI were calculated for each variable. Risk of RILD (NTCP value) for particular patient can be estimated using the following equation:

$$NTCP = \frac{1}{1+e^{-S(x)}} \quad \text{, in which}$$

$$S(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

where x is the prognostic variable, β_0 is a constant value and β_i $(1 \leq i \leq n)$ represented the regression coefficients for i^{th} covariate (x).

Model performance was determined by discriminative ability using the AUC and by assessing agreement between predicted and observed outcomes using the calibration plot. Further, a spike plot of the distribution of events and non-events was displayed, as well as a lowess smoother allowing assessment of the calibration at the individual patient level. Hosmer-Lemeshow goodness-of-fit test was further used to evaluate the fitted NTCP model. We also performed model validation in the test dataset and compared with LKB NTCP models from previous studies. Statistical analyses were conducted in IBM SPSS statistics (version 22.0, SPSS, Inc., Chicago, IL) and STATA LP (version 13.1, StataCorp, College Station, TX).

Results

Patient and treatment outcomes

The training dataset consisted 179 HCC and 101 ICC. Majority were male (73.2%) and CP-A classification (72.1%). A half of patients had positive chronic viral hepatitis infection status. The median prescription dose was 45 Gy in a daily fraction of 1.8-10 Gy. Pretreatment characteristics are listed in Table 10. The incidence of RILD in overall patients was 31.8% (89/280): 10.6%, 43.8%, 28.6% and 60% in CP-A, non-hepatitis (12/113), CP-A, hepatitis (39/89), CP-B/C, non-hepatitis (8/28) and CP-B/C, hepatitis (30/50), respectively (p<0.001).

With the median follow-up time of 9.5 months, median OS for overall patients was 13.3 months (95%CI 10.7-15.8 months), 17.9 months (95%CI 13.8-22 months) *versus* 4.8 months (95%CI 3.5-6 months) in patients with non-RILD *versus* RILD, respectively (p<0.001).

Table 10 Pretreatment characteristics of model training dataset and test dataset

	Training dataset (N=280)	**Test dataset (N=46)**
Age (years), mean (SD)	60.3 (12.3)	59.6 (9.1)
Gender		
male	205 (73.2%)	38 (82.6%)
female	75 (26.8%)	8 (17.4%)
Primary disease		
hepatocellular carcinoma	179 (63.9%)	39 (84.8%)
intrahepatic cholangiocarcinoma	101 (36.1%)	7 (15.2%)
Presence of portal vein thrombosis	137 (48.9%)	31 (67.4%)
T-classification		
early T stage (T2 or less)	85 (30.4%)	14 (30.4%)
late T stage (T3 or more)	195 (69.6%)	32 (69.6%)
N-classification		
no lymph node involvement (N0)	210 (75%)	41 (89.1%)
positive lymph node involvement (N1)	70 (25%)	5 (10.9%)
Child-Pugh classification		
A	202 (72.1%)	38 (82.6%)
B or C	78 (27.9%)	8 (17.4%)
Viral hepatitis B or C infection	139 (49.6%)	32 (69.6%)
Previous treatments		
surgery	76 (27.1%)	8 (17.4%)
transarterial chemoembolization	113 (40.4%)	29 (63%)
chemotherapy	42 (15%)	4 (8.7%)
Radiation technique		
3D-CRT	124 (44.3%)	5 (10.9%)
IMRT/VMAT	106 (37.9%)	33 (71.7%)
SBRT	50 (17.8%)	8 (17.4%)

	Training dataset (N=280)	Test dataset (N=46)
Median (IQR) total dose (Gy)	45 (30-50)	45 (40-50)
Median (IQR) Number of fraction (fractions)	10 (10-25)	10 (5-15)
Median (IQR) Dose per fraction (Gy)	3 (1.8-4)	3 (3-8)
Mean (SD) normal liver volume	1197.3 (452.3)	1490.3 (596.6)
Mean (SD) liver dose (Gy)	17.7 (7.8)	17.8 (8.4)
Effective volume (SD) (cc)	0.4 (0.5)	0.5 (0.2)
$V_{5\ Gy}$ (SD) (%)	67.0 (23.9)	66.6 (22.8)
$V_{10\ Gy}$ (SD) (%)	55.3 (25.1)	51.4 (20.1)
$V_{20\ Gy}$ (SD) (%)	41.6 (23.5)	31.3 (15.5)
$V_{30\ Gy}$ (SD) (%)	31.7 (21.5)	21.5 (12.8)
$V_{40\ Gy}$ (SD) (%)	25.1 (18.6)	12.9 (9.7)

Abbreviation: 3D-CRT=three dimension conformal radiotherapy, IMRT=intensity modulated radiotherapy, VMAT=volumetric modulated arc therapy, SBRT=stereotactic body radiotherapy, IQR=interquatile range, SD=standard deviation

In the univariate analysis, diagnosis of HCC, presence of PVT, late T stage, CP-B or C classification, positive hepatitis status, prior surgery and TACE, and MLD were significantly associated with RILD occurrence as shown in Table 11. Notably, when stratified patients according to primary disease (HCC *versus* ICC), there were more patients with good prognosticators in ICC group. For example, rates of CP-A and negative hepatitis were >80% in ICC patients, but only 67% and 30% in HCC patients (p=0.011 for CP classification and p<0.001 for hepatitis).

Table 11 Univariate logistic regression analysis for radiation-induced liver disease

Variables		ORs	95%CI		*P-value*
			Lower	Upper	
Age (year)		0.990	0.970	1.010	*0.323*
Gender	female	1.00			
	male	1.53	0.84	2.77	*0.162*

Variables		ORs	95%CI		*P-value*
			Lower	**Upper**	
Diagnosis	intrahepatic cholangiocarcinoma	1.00			
	hepatocellular carcinoma	9.61	4.41	20.96	*<0.001*
Presence of portal vein thrombosis		4.85	2.78	8.47	*<0.001*
T-classification	early T stage	1.00			
	late T stage	2.85	1.52	5.35	*0.001*
N-classification	N0	1.00			
	N1	0.68	0.37	1.24	*0.209*
Child-Pugh classification	A	1.00			
	B or C	2.81	1.63	4.85	*<0.001*
Viral hepatitis infection status	negative	1.00			
	positive	5.96	3.34	10.63	*<0.001*
Surgery		0.15	0.07	0.34	*<0.001*
TACE		1.85	1.11	3.08	*0.018*
Chemotherapy		0.54	0.25	1.18	*0.122*
Radiation technique	3D-conformal radiotherapy	1.00			
	intensity modulated radiotherapy	1.111	0.545	2.266	*0.084*
	stereotactic body radiotherapy	1.102	0.531	2.285	*0.068*
Normal liver volume (cc)		1.00	1.00	1.00	*0.670*
Mean liver dose (Gy)		1.04	1.00	1.07	*0.030*
$V_{5\ Gy}$ (%)		1.01	1.00	1.02	*0.157*
$V_{10\ Gy}$ (%)		1.00	0.99	1.01	*0.641*
$V_{20\ Gy}$ (%)		1.00	0.99	1.01	*0.790*
$V_{30\ Gy}$ (%)		1.00	0.99	1.01	*0.631*
$V_{40\ Gy}$ (%)		1.00	0.99	1.02	*0.592*

Abbreviation: OR=odds ratio, 95%CI=95% confidence interval, $V_{x\ Gy}$=volume of normal liver receiving x Gy

NTCP model development

In the multivariate logistic regression analysis with 1000 bootstrapping, the significant variables in the final model included diagnosis of HCC, CP-B/C, positive hepatitis, and MLD. The OR, regression coefficients and 95%CI of each selected variable were shown in Table 12. The NTCP value for the individual patient can be calculated using the equation:

$$NTCP = \frac{1}{1+e^{-S(x)}}$$, in which

$S(x)$ = -3.94 + (diagnosis of HCC * 1.74) + (CP-B or C * 0.90) + (viral hepatitis infection status * 1.06) + (MLD* 0.05)

Table 12 Multivariable logistic regression analysis for radiation-induced liver disease

Variables	ORs	95%CI		p-value	Regression coefficients	95%CI		Multiplication value
		Lower	Upper			Lower	Upper	
Diagnosis								
ICC	1.00							
HCC	5.68	2.22	14.52	<0.001	1.74	0.81	2.67	0=no,1=yes
Child-Pugh classification								
A	1.00							
B or C	2.47	1.28	4.78	0.007	0.90	0.27	1.54	0=no, 1=yes
Viral hepatitis status								
negative	1.00							
positive	2.89	1.53	5.46	0.001	1.06	0.44	1.68	0=no, 1=yes
Mean liver dose	1.06	1.01	1.10	0.013	0.05	0.01	0.09	Dose in Gy
Constant					-3.94	-5.25	-2.63	

Abbreviation: OR=odds ratio, 95%CI=95% confidence interval, ICC=intrahepatic cholangiocarcinoma, HCC=hepatocellular carcinoma; CP=Child-Pugh classification

Mean NTCP values for patients with RILD (n=89) and non-RILD (n=191) were 49.1% (95%CI 44.8%-53.4%) and 23.7% (95%CI 20.9%-26.6%), respectively. The model had high specificity (86%) and negative predictive value (81%). AUC for this 4-variable model was 0.81 (95%CI 0.75-0.86) indicating good model performance. The HL test also showed an agreement between predicted and observed outcomes (χ^2 =6.57, p=0.58) as demonstrated by calibration plot in Figure 13.

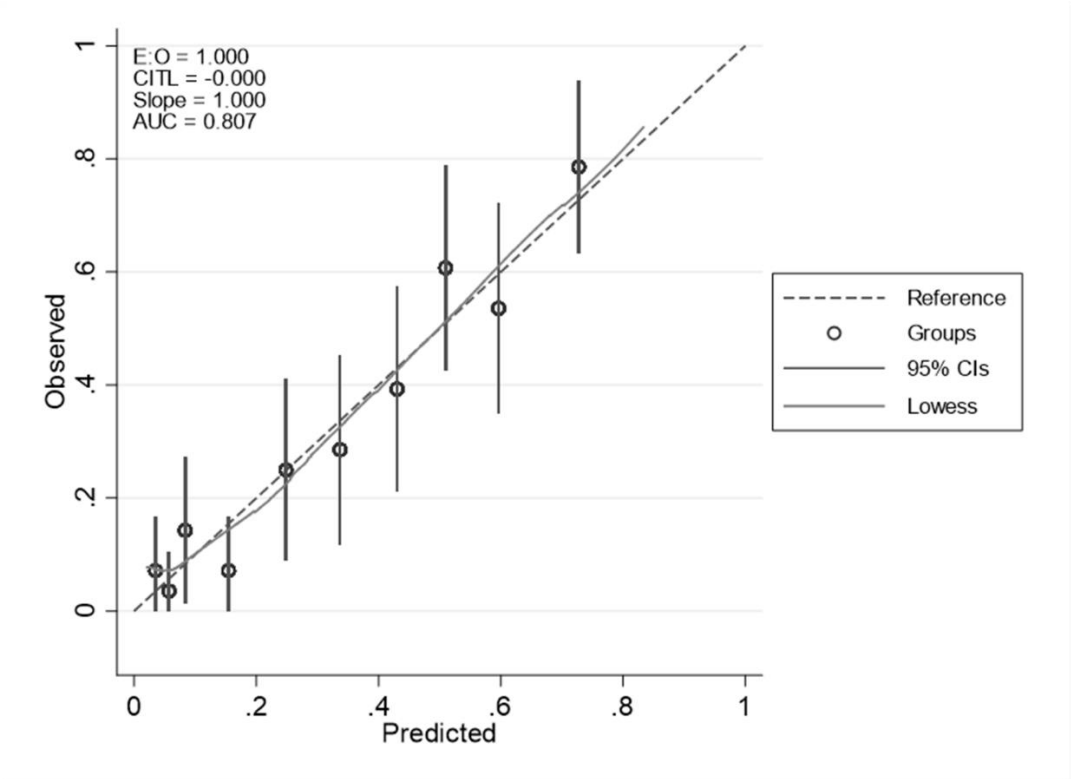


Figure 13 Calibration plot for the predictive model for RILD. The solid line represents the calibration graph of the model, the dash line represents the identical line, and the circle with error bar represents the Hosmer-Lemeshow groups.

According to 3 clinical parameters, patients were stratified into 8 subgroups. Baseline NTCP values for each subgroup, without consideration of MLD (MLD=0 Gy), were estimated as shown in Table 13.

Table 13 Estimated baseline NTCP values for radiation-induced liver disease considering pretreatment clinical characteristics

Diagnosis	Child-Pugh classification	Viral hepatitis Infection status	Baseline estimated NTCP for RILD
ICC	A	negative	1.9%
		positive	5.3%
	B or C	negative	4.6%
		positive	12.1%
HCC	A	negative	10.0%
		positive	24.2%
	B or C	negative	21.4%
		positive	44.0%

Abbreviration: ICC=intrahepatic cholagiocarcinoma; HCC=hepatocellular carcinoma

Model validation

Demographic patient and tumor data of the test dataset are listed in Table 10, where more patients with HCC and positive hepatitis were observed. Of the 46 patients, RILD occurred in 21 patients (45.7%). When we applied data of the test dataset in our multivariable NTCP model, mean NTCP values were 51.7% (95%CI 44.4%-58.9%) for RILD patients and 25.9% (95%CI 18.9-32.6%) for non-RILD group. The model accuracy and specificity were 74% and 97%, respectively. Overall model performance was good for our model with the AUC of 0.86 (95%CI 0.75-0.96).

In comparison of our multivariable NTCP model with LKB NTCP models from previous studies, Table 14 summarizes characteristics of patients used to derive NTCP model in each study. With their LKB NTCP parameters, mean NTCP values of the test dataset were calculated. AUCs of other studies were considerably dropped to <0.70. Noting that, prior to NTCP calculation, MLD was calculated using f_s and n according to each study protocol.

Table 14 Comparisons of NTCP values and model performance in test dataset (N=46) using LKB NTCP parameters from previous studies or multivariable logistic regression coefficients in this study

Studies						LKB NTCP parameters				$NTCP_{test}$		AUC in
Year	**N (RILD)**	**Dx**	**CP**	**Hep**	**Fx (Gy)**	**group**	**TD_{50}**	**m**	**n**	**Non-RILD**	**RILD**	**test dataset**
(Lawrence et al., 1992)	79 (9)	PLC+ LM	A	NA	1.5-1.65*	-	40	0.15	0.69	0.2%	7.8%	0.66 (0.49-0.82)
(Dawson et al., 2002)	203 (19)	PLC+ LM	A	NA	1.5-1.65*	-	39.8	0.12	0.97	6.2%	10.4%	0.65 (0.48-0.81)
(Cheng et al., 2004)	89 (17)	HCC	68 A/ 21 B	65 +/ 24 -	1.8-3	-	49.4	0.39	0.35	17.7%	22.9%	0.59 (0.42-0.76)
(Xu et al., 2006)	109 (17)	PLC	93 A/ 16 B	86 +/ 23 -	4-6	CP-A	40.5	0.28	1.1	0.2%	0.2%	0.67 (0.50-0.83)
						CP-B	23	0.43	0.7	14.3%	22.9%	0.62 (0.46-0.79)
(Tai et al., 2009)	414 (57)	PLC	336 A/ 78 B	NA	1.5-1.65* 1.8-2 4.6	CP-A	40.3	0.36	1	7.9%	17.3%	0.66 (0.49-0.82)
						CP-B	23.9	0.41	1	8.2%	15.6%	0.66 (0.49-0.82)
This study	280 (89)	PLC	202 A/ 78 B	139 +/ 141 -	1.8-10	-	-	-	-	25.9%	51.7%	0.86 (0.75-0.96)

* twice-daily radiotherapy

Abbreviation: PLC=primary liver cancer; LM=liver metastases; HCC=hepatocellular carcinoma; CP=Child-Pugh classification; Hep=viral hepatitis infection status; TD=total prescription dose; Fx=dose fractionation; MLD=mean liver dose; RILD=radiation-induced liver disease; LKB=Lyman-Kutcher Burman; NTCP=normal tissue complication probability; $NTCP_{test}$=NTCP value in test dataset (N=46) using LKB NTCP parameters from previous studies or multivariable logistic regression coefficients in this study; AUC=area under the curve; TD_{50}=tolerance dose of whole liver with 50% probability of developing toxicity; m=slope at TD_{50}; n=volume effect; NA=not applicable

Discussion

In this study, we introduced the prediction model to identify primary liver cancer patients at risk of developing RILD by integrating clinical and dosimetric parameters. Using multivariate logistic regression analysis, the final model consisted of 4 significant parameters including diagnosis (HCC *versus* ICC), pretreatment CP classification (CP-A *versus* CP-B/C), HBV/HCV infection status (positive *versus* negative) and MLD. Model performance and validity were excellent compared to the preexisting models.

The incidence of RILD ranged from 9.4%-19.1% in previous reports (Cheng et al., 2005, Cheng et al., 2004, Dawson et al., 2002, Kim et al., 2007, Liang et al., 2006, Xu et al., 2006). High RILD incidence in this study (33%) might be due to greater intrinsic biosusceptibility characteristics in our patient population. Additionally, in contrast to ICC which developed from bile duct epithelial cells, HCC originated from hepatocytes which frequently associated with chronic viral hepatitis and liver cirrhosis, especially in Asian population (Cheng et al., 2002, Kim et al., 2007, Liang et al., 2006). Thus, some hepatocytes were conceivably infected and revealed some degree of cirrhosis. Correspondingly, we demonstrated that HCC diagnosis was the most significant predictor with adjusted OR of 5.35, indicating higher radiosensitivity of liver tissue compared with ICC. The other reason of the high incidence rate of RILD in our study was the use of liver function-specific RILD criteria, i.e., classic RILD for CP-A, non-hepatitis patients and nonclassic RILD for those with CP-B/C or hepatitis.

The predictive factors included in our model were mostly in line with previous studies (Cheng et al., 2005, Cheng et al., 2004, Dawson et al., 2002, Kim et al., 2007, Liang et al., 2006, Seong et al., 2003, Tai et al., 2009, Xu et al., 2006). However, some worse prognostic factors (PVT, late T stage and TACE) were statistically significant in univariate analysis as previously reported (Liang, 2006;Seong, 2003) but were eventually excluded from the final model due to inadequately significant effect. Unlike earlier study (Dawson, 2002), chemotherapy was not associated with increased RILD risk in this report (OR=0.54, p=0.122). The plausible explanation was that we usually administered chemotherapy as an adjuvant treatment following definitive surgery for ICC, in contrast to the concomitant use with twice-daily radiotherapy in a previous study (Dawson, 2002).

As for the test dataset, model performance of our 4-parameter NTCP model was apparently superior to the preexisting LKB NTCP models. Some reasons were unavoidable similarity of our training and test datasets considering biological and treatment aspects, and capability of concerning many risk factors at a time instead of only dosimetric parameter. For example, Michigan study composed of patients with normal liver function, but unknown viral hepatitis status. Besides, their treatment regimen was twice daily radiotherapy concurrent with chemotherapy (Dawson, 2002). Although CP-B and HBV carrier status were taken into account in Asian studies (Cheng et al., 2004, Tai et al., 2009, Xu et al., 2006), but NTCP parameters which were derived from a small-size subgroup limited the application of the models.

With our multivariable NTCP model, treatment strategies can be modified in an individualized basis. In advanced head and neck cancers, Jakobi *et al.* evaluated the feasibility for dose-escalation schedule by assessing tumor control probability and NTCP and found approximately 10% increase in TCP with <2% increase in NTCP (Jakobi et al., 2015b). In addition, NTCP model-based treatment plan optimization is clinically achievable. Kierkels *et al*. optimized treatment plans using multivariable NTCP-based objective functions for tube feeding dependence and/or dysphagia to obtain lower dose distribution to organs-at-risk at comparable dose to the tumor (Kierkels et al., 2016). According to Table 13, patients with only 1 clinical risk factor had baseline RILD risk ≤10%, thus tumor dose-escalation for unresectable diseases might be possible. For patients with 2 clinical risks, treatment plan should be optimized with high respect to MLD. And, for those with all 3 risk features, radiotherapy might not be appropriate due to high baseline NTCP value (44%).

With attempts to improve tumor control and minimize NTCP, many modern radiotherapy techniques have been developed, particularly PBT but with high expense. The NTCP-model based approach is a promising method for selecting patients for PBT due to some limitations of conducting RCT. This approach has been approved as an alternative high-level evidence for PBT implementation by the health council and health care insurance board of the Netherlands (Lambin et al., 2013, Langendijk et al., 2013) and as a guide for creating an enriched cohorts of patients for randomized study between XRT *versus* PBT (Widder et al., 2016). For liver tumors, studies from Japan proposed the implementation of this approach to select patients for PBT (Toramatsu et al., 2013, Kobashi et al., 2018, Shirato et al., 2018) including our study in Chapter 2. The NTCP

model-based approach for the selective use of PBT in primary liver cancer patients will be further studied using our multivariable NTCP model.

The limitation of this study was retrospective data collection. Heterogeneity of disease characteristics and treatment techniques/regimens inevitably affected the predictive power of model (Beetz et al., 2012b). Therefore, our model needs external validation, especially in different clinical entity, prior to clinical application. Besides, the novel predictive markers including circulating inflammatory proteins and functional imaging as well as radiogenomic data have been of greater interest for RILD (Cao et al., 2008, Schaue et al., 2012, El Naqa et al., 2018). These potential biomarkers are capable of being integrated into the NTCP model for improving prediction ability and further implementing in treatment plan optimization.

To the best of our knowledge, the current study is the first to develop multivariable NTCP model combining clinical and dosimetric factors for predicting the risk of RILD using relatively large patient cohort. Moreover, FED provided the generalized application of this model among various treatment schedules and fractionations. Further study will be focusing on multivariable NTCP model-based approach considering uncertainties of NTCP and ΔNTCP which might improve the reliability of the model prediction in the general population

Conclusions

・In primary liver cancers, radiation therapy is effective but can induce radiation-induced liver disease (RILD). Normal tissue complication probability (NTCP) model-based approach has been expected to predict the risk of RILD in each patient especially in the selection of patients for proton beam therapy (PBT) rather than x-ray treatment (XRT). Precise investigation about the reliability of the NTCP model-based approach of RILD is now mandatory in the world because the number of PBT centers is limited.

・We found that mathematical estimation of uncertainties of NTCP and the difference of NTCP between two radiotherapy methods (ΔNTCP) is possible to improve the reliability of the NTCP model-based approach in the prediction of RILD.

・We found that delta method is useful to describe variance and covariance of LKB NTCP model parameters to assess uncertainties of NTCP and ΔNTCP using the real-world dataset of 239 patients as well as the dataset derived from previous publication.

・We have found appropriate parameters for the dose-response between mean dose to liver and the risk of RILD based on LKB NTCP model for subgroups of patients stratified by biosusceptibility factors (CP classification and hepatitis status). We established the dose-response curves with 68% confidence interval using the NTCP model for each subgroup and found that the parameters markedly varied among subgroups. Overall model performance was acceptable with AUC of 0.714 (95%CI, 0.650-0.779). The AUC was favorable when we compared it with the AUC of 0.569 (95%CI, 0.496-0.642) which was obtained with the Michigan's common parameters for RILD.

・The values of NTCP and ΔNTCP derived from the subgroup-specific LKB NTCP models markedly varied among subgroups. We obtained confidence interval of ΔNTCP using delta method and showed its potential to improve the reliability of the NTCP model-based approach.

・We first established the scientific background about the possible use of ΔNTCP with confidence interval in this study. Following to that, we first proposed to use ΔNTCP with confidence interval in the selection of an appropriate external-beam radiotherapy. It was more conservative than the use of central estimate of ΔNTCP thus avoid overuse of PBT, and increased confidence for patient selection. These findings underlined the importance of appropriate patient selection and continuous improvement to find the proper subgrouping.

- The shortcoming of the subgroup-specific LKB NTCP models is the way to clustering. Apart from the CP classification and hepatitis status, we integrated other clinical and dosimetric parameters to develop a novel NTCP model based on multivariate logistic regression.
- Using multivariate logistic regression analysis, the final model consisted of 4 significant parameters including 3 clinical parameters (primary diagnosis, CP classification, and hepatitis status) and 1 dosimetric parameter, MLD.
- Model performance and validity of the novel multivariate NTCP model were excellent compared to the preexisting models. Overall model performance of our multivariate NTCP model was good with the AUC of 0.86 (95%CI 0.75-0.96). AUCs of NTCP models which were previous published were considerably lower than our multivariate NTCP model (<0.70).
- However, we have yet achieved to establish a dose-response curve between MLD and the NTCP value with confidence interval using the multivariate NTCP model.
- Further study should be focusing on uncertainties of NTCP and ΔNTCP using the multivariable NTCP model which we established in this study. We hope that it should further improve the reliability of the NTCP model-based approach.
- Since the multivariable NTCP model will fit well with machine learning and big data science, further study should be able to use more sophisticated statistical analytics.

Acknowledgement

Firstly, I would like to express my sincere gratitude to Professor Hiroki Shirato, the head of Department of Radiation Medicine, Faculty of Medicine, Hokkaido University; Dean of the Graduate School of Biomedical Science and Engineering; and Director of Global Station for Quantum Medical Science and Engineering (GSQ), Global Institution for Collaborative Research and Education (GI-CoRE), and Professor Shinichi Shimizu, the head of Department of Radiation Oncology, Faculty of Medicine, Hokkaido University, for giving me the opportunity of this study and for their strongly supports on my research. They provide me thoughtful idea and comments on my research. Without their guidance for formulating research topic, research methodology and discussion, this research would not have been possible.

Secondly, I deeply appreciate the great help from Associate Professor Keiji Kobashi who has been supporting me from the beginning of my research. He gives me invaluable help on understanding of the topic, complicated computational supports and fruitful comments for discussion and further studies.

Thirdly, I would like to acknowledge Professor Yoichi Ito and Dr.Natchalee Srimaneekarn whose expertise is very helpful for statistical analysis. I have received a great deal of support and assistance from them.

Finally, I would like to offer my special thanks to my colleges and staffs in Hokkaido University, Japan and in Chulalongkorn University, Thailand for their supports.

Disclosure of Conflict of Interest

The author declares no conflict of interest.

References

Armitage, P., Berry, G., Matthews, J. (2001). Statistical methods in medical research (United Kingdom: Wiley-Blackwell).

Beetz, I., Schilstra, C., Burlage, F.R., Koken, P.W., Doornaert, P., Bijl, H.P., Chouvalova, O., Leemans, C.R., De Bock, G.H., Christianen, M.E., et al. (2012A). Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. Radiother. Oncol. *105***,** 86-93.

Beetz, I., Schilstra, C., Van Luijk, P., Christianen, M.E., Doornaert, P., Bijl, H.P., Chouvalova, O., Van Den Heuvel, E.R., Steenbakkers, R.J., Langendijk, J.A. (2012B). External validation of three dimensional conformal radiotherapy based NTCP models for patient-rated xerostomia and sticky saliva among patients treated with intensity modulated radiotherapy. Radiother. Oncol. *105***,** 94-100.

Bijman, R.G., Breedveld, S. (2017). Impact of model and dose uncertainty on model-based selection of oropharyngeal cancer patients for proton therapy. Acta Oncol. *56***,** 1444-1450.

Blanchard, P., Wong, A.J., Gunn, G.B., Garden, A.S., Mohamed, A.S.R., Rosenthal, D.I., Crutison, J., Wu, R., Zhang, X., Zhu, X.R., et al. (2016). 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. *121***,** 381-386.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A. (2018). Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. *68***,** 394-424.

Bush, D.A., Kayali, Z., Grove, R., Slater, J.D. (2011). The safety and efficacy of high-dose proton beam radiotherapy for hepatocellular carcinoma: A phase 2 prospective trial. Cancer *117***,** 3053-3059.

Bush, D.A., Smith, J.C., Slater, J.D., Volk, M.L., Reeves, M.E., Cheng, J., Grove, R., De Vera, M.E. (2016). Randomized clinical trial comparing proton beam radiation therapy with transarterial chemoembolization for hepatocellular carcinoma: Results of an interim analysis. Int. J. Radiat. Oncol. Biol. Phys. *95***,** 477-482.

Cao, Y., Pan, C., Balter, J.M., Platt, J.F., Francis, I.R., Knol, J.A., Normolle, D., Ben-Josef, E., Ten Haken, R.K., Lawrence, T.S. (2008). Liver function after irradiation based on computed tomographic portal vein perfusion imaging. Int. J. Radiat. Oncol. Biol. Phys. *70*, 154-160.

Cheng, J.C., Wu, J.K., Huang, C.M., Liu, H.S., Huang, D.Y., Cheng, S.H., Tsai, S.Y., Jian, J.J., Lin, Y.M., Cheng, T.I., et al. (2002). Radiation-induced liver disease after three-dimensional conformal radiotherapy for patients with hepatocellular carcinoma: dosimetric analysis and implication. Int. J. Radiat. Oncol. Biol. Phys. *54*, 156-162.

Cheng, J.C., Wu, J.K., Lee, P.C., Liu, H.S., Jian, J.J., Lin, Y.M., Sung, J.L., Jan, G.J. (2004). Biologic susceptibility of hepatocellular carcinoma patients treated with radiotherapy to radiation-induced liver disease. Int. J. Radiat. Oncol. Biol. Phys. *60*, 1502-9.

Cheng, J.C., Liu, H.S., Wu, J.K., Chung, H.W., Jan, G.J. (2005). Inclusion of biological factors in parallel-architecture normal-tissue complication probability model for radiation-induced liver disease. Int. J. Radiat. Oncol. Biol. Phys. *62*, 1150-1156.

Cheng, Q., Roelofs, E., Ramaekers, B.L., Eekers, D., Van Soest, J., Lustberg, T., Hendriks, T., Hoebers, F., Van Der Laan, H.P., Korevaar, E.W., et al. (2016). 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. *118*, 281-285.

Chiba, T., Tokuuye, K., Matsuzaki, Y., Sugahara, S., Chuganji, Y., Kagei, K., Shoda, J., Hata, M., Abei, M., Igaki, H., et al. (2005). Proton beam therapy for hepatocellular carcinoma: A retrospective review of 162 patients. Clin. Cancer Res. *11*, 3799-3805.

Dawson, L.A. (2009). Protons or photons for hepatocellular carcinoma? Let's move forward together. Int. J. Radiat. Oncol. Biol. Phys. *74*, 661-663.

Dawson, L.A., Normolle, D., Balter, J.M., Mcginn, C.J., Lawrence, T.S., Ten Haken, R.K. (2002). Analysis of radiation-induced liver disease using the Lyman NTCP model. Int. J. Radiat. Oncol. Biol. Phys. *53*, 810-821.

Durante, M. (2015). Charged particles for liver cancer. Ann. Transl. Med. *3*, 363.

Durante, M., Orecchia, R., Loeffler, J.S. (2017). Charged-particle therapy in cancer: clinical uses and future perspectives. Nat. Rev. Clin. Oncol. *14*, 483-495.

Efron, B., Tibshirani, R. (1993). An introduction to the bootstrap (Boca Raton, Fl: Chapman & Hall).

El Naqa, I., Bradley, J., Blanco, A.I., Lindsay, P.E., Vicic, M., Hope, A., Deasy, J.O. (2006). Multivariable modeling of radiotherapy outcomes, including dose-volume and clinical factors. Int. J. Radiat. Oncol. Biol. Phys. *64***,** 1275-1286.

El Naqa, I., Johansson, A., Owen, D., Cuneo, K., Cao, Y., Matuszak, M., Bazzi, L., Lawrence, T.S., Ten Haken, R. K. (2018). Modeling of normal tissue complications using imaging and biomarkers after radiation therapy for hepatocellular carcinoma. Int. J. Radiat. Oncol. Biol. Phys. *100***,** 335-343.

Emami, B., Lyman, J., Brown, A., Coia, L., Goitein, M., Munzenrider, J.E., Shank, B., Solin, L. J., Wesson, M. (1991). Tolerance of normal tissue to therapeutic irradiation. Int. J. Radiat. Oncol. Biol. Phys. *21***,** 109-122.

Fukumitsu, N., Sugahara, S., Nakayama, H., Fukuda, K., Mizumoto, M., Abei, M., Shoda, J., Thono, E., Tsuboi, K., Tokuuye, K. (2009). A prospective study of hypofractionated proton beam therapy for patients with hepatocellular carcinoma. Int. J. Radiat. Oncol. Biol. Phys. *74***,** 831-836.

Goitein, M., Cox, J. D. (2008). Should Randomized Clinical Trials be required for proton radiotherapy? J. Clin. Oncol. *26***,** 175-176.

Granovetter, M. (2016). Proton radiotherapy for primary liver cancers. Lancet Oncol. *17***,** E49.

Grau, C. (2013). The model-based approach to clinical studies in particle radiotherapy--A new concept in evidence based radiation oncology? Radiother. Oncol. *107***,** 265-266.

Hellman, S., Hellman, D. S. (1991). Of mice but not men. Problems of the Randomized Clinical Trial. N. Engl. J. Med. *324***,** 1585-1589.

Hong, T.S., Wo, J.Y., Yeap, B.Y., Ben-Josef, E., Mcdonnell, E.I., Blaszkowsky, L.S., Kwak, E.L., Allen, J.N., Clark, J.W., Goyal, L., et al. (2016). 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. *34***,** 460-468.

Igaki, H., Mizumoto, M., Okumura, T., Hasegawa, K., Kokudo, N., Sakurai, H. (2018). A systematic review of publications on charged particle therapy for hepatocellular carcinoma. Int J. Clin. Oncol. *23***,** 423-433.

Jakobi, A., Bandurska-Luque, A., Stutzer, K., Haase, R., Lock, S., Wack, L. J., Monnich, D., Thorwarth, D., Perez, D., Luhr, A., et al. (2015A). 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. *92***,** 1165-1174.

Jakobi, A., Luhr, A., Stutzer, K., Bandurska-Luque, A., Lock, S., Krause, M., Baumann, M., Perrin, R., Richter, C. (2015B). Increase in tumor control and normal tissue complication probabilities in advanced head-and-neck cancer for dose-escalated intensity-modulated photon and proton therapy. Front. Oncol. *5***,** 256.

Jakobi, A., Stutzer, K., Bandurska-Luque, A., Lock, S., Haase, R., Wack, L.J., Monnich, D., Thorwarth, D., Perez, D., Luhr, A., et al. (2015C). NTCP reduction for advanced head and neck cancer patients using proton therapy for complete or sequential boost treatment versus photon therapy. Acta Oncol. *54***,** 1658-1664.

Kawashima, M., Furuse, J., Nishio, T., Konishi, M., Ishii, H., Kinoshita, T., Nagase, M., Nihei, K., Ogino, T. (2005). Phase II study of radiotherapy employing proton beam for hepatocellular carcinoma. J. Clin. Oncol. *23***,** 1839-1846.

Kierkels, R.G.J., Wopken, K., Visser, R., Korevaar, E.W., Van Der Schaaf, A., Bijl, H.P., Langendijk, J.A. (2016). Multivariable normal tissue complication probability model-based treatment plan optimization for grade 2-4 dysphagia and tube feeding dependence in head and neck radiotherapy. Radiother. Oncol. *121***,** 374-380.

Kim, T.H., Kim, D.Y., Park, J.W., Kim, S.H., Choi, J.I., Kim, H.B., Lee, W.J., Park, S.J., Hong, E.K., Kim, C. M. (2007). Dose-volumetric parameters predicting radiation-induced hepatic toxicity in unresectable hepatocellular carcinoma patients treated with three-dimensional conformal radiotherapy. Int. J. Radiat. Oncol. Biol. Phys. *67***,** 225-231.

Kim, T.H., Park, J.W., Kim, Y.J., Kim, B.H., Woo, S.M., Moon, S.H., Kim, S.S., Koh, Y.H., Lee, W.J., Park, S.J., et al. (2015). Phase I dose-escalation study of proton beam therapy for inoperable hepatocellular carcinoma. Cancer Res. Treat. *47***,** 34-45.

Kobashi, K., Prayongrat, A., Kimoto, T., Toramatsu, C., Dekura, Y., Katoh, N., Shimizu, S., Ito, Y. M., Shirato, H. (2018). Assessing the uncertainty in a normal tissue complication probability difference (NTCP): radiation-induced liver disease (RILD) in liver tumour patients treated with proton vs x-ray therapy. J. Radiat. Res. *59***,** I50-I57.

Kutcher, G.J., Burman, C. (1989). Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method. Int. J. Radiat. Oncol. Biol. Phys. *16***,** 1623-1630.

Kutcher, G.J., Burman, C., Brewster, L., Goitein, M., Mohan, R. (1991). Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations. Int. J. Radiat. Oncol. Biol. Phys. *21***,** 137-146.

Lambin, P., Van Stiphout, R. G., Starmans, M. H., Rios-Velazquez, E., Nalbantov, G., Aerts, H. J., Roelofs, E., Van Elmpt, W., Boutros, P.C., Granone, P., et al. (2013). Predicting outcomes in radiation oncology--Multifactorial decision support systems. Nat. Rev. Clin. Oncol. *10***,** 27-40.

Langendijk, J.A., Lambin, P., De Ruysscher, D., Widder, J., Bos, M., Verheij, M. (2013). Selection of patients for radiotherapy with protons aiming at reduction of side effects: The Model-Based Approach. Radiother. Oncol. *107***,** 267-273.

Lawrence, T.S., Ten Haken, R.K., Kessler, M.L., Robertson, J.M., Lyman, J.T., Lavigne, M.L., Brown, M.B., Duross, D.J., Andrews, J.C., Ensminger, W.D. et al. (1992). The use of 3-D dose volume analysis to predict radiation hepatitis. Int. J. Radiat. Oncol. Biol. Phys. *23***,** 781-788.

Lee, S.U., Park, J.W., Kim, T.H., Kim, Y.J., Woo, S.M., Koh, Y.H., Lee, W.J., Park, S.J., Kim, D.Y., Kim, C. M. (2014). Effectiveness and safety of proton beam therapy for advanced hepatocellular carcinoma with portal vein tumor thrombosis. Strahlenther. Onkol. *190***,** 806-814.

Liang, S.X., Zhu, X.D., Xu, Z.Y., Zhu, J., Zhao, J.D., Lu, H.J., Yang, Y.L., Chen, L., Wang, A. Y., Fu, X.L., et al. (2006). Radiation-induced liver disease in three-dimensional conformal radiation therapy for primary liver carcinoma: the risk factors and hepatic radiation tolerance. Int. J. Radiat. Oncol. Biol. Phys. *65***,** 426-434.

Luxton, G., Keall, P.J., King, C.R. (2008). A new formula for normal tissue complication probability (NTCP) as a function of equivalent uniform dose (EUD). Phys. in Med. Biol., *53***,** 23-36.

Lyman, J.T. (1985). Complication probability as assessed from dose-volume histograms. Radiat. Res. *Suppl, 8***,** S13-S19.

Marks, L.B., Yorke, E.D., Jackson, A., Ten Haken, R.K., Constine, L.S., Eisbruch, A., Bentzen, S.M., Nam, J., Deasy, J.O. (2010). Use of normal tissue complication probability models in the clinic. Int. J. Radiat. Oncol. Biol. Phys. *76***,** S10-S19.

Martin, C., Brad, O., Kerwyn, F., Annette, H., Sarah, G., Martin, E. (2014). An MLE method for finding LKB NTCP model parameters using Monte Carlo uncertainty estimates. J. Phys. *489***,** 012087.

Mizumoto, M., Tokuuye, K., Sugahara, S., Nakayama, H., Fukumitsu, N., Ohara, K., Abei, M., Shoda, J., Tohno, E., Minami, M. (2008). Proton beam therapy for hepatocellular carcinoma adjacent to the porta hepatis. Int. J. Radiat. Oncol. Biol. Phys. *71***,** 462-467.

Nakayama, H., Sugahara, S., Tokita, M., Fukuda, K., Mizumoto, M., Abei, M., Shoda, J., Sakurai, H., Tsuboi, K., Tokuuye, K. (2009). Proton beam therapy for hepatocellular carcinoma: The university of Tsukuba experience. Cancer *115***,** 5499-5506.

Nakayama, H., Sugahara, S., Fukuda, K., Abei, M., Shoda, J., Sakurai, H., Tsuboi, K., Matsuzaki, Y., Tokuuye, K. (2011). Proton beam therapy for hepatocellular carcinoma located adjacent to the alimentary tract. Int. J. Radiat. Oncol. Biol. Phys. *80***,** 992-995.

Niemierko, A. (1997). Reporting and analyzing dose distributions: A concept of equivalent uniform dose. Med. Phys. *24***,** 103-110.

Osmundson, E.C., Wu, Y., Luxton, G., Bazan, J.G., Koong, A.C., Chang, D.T. (2015). Predictors of toxicity associated with stereotactic body radiation therapy to the central hepatobiliary tract. Int. J. Radiat. Oncol. Biol. Phys. *91***,** 986-994.

Pan, C.C., Kavanagh, B.D., Dawson, L.A., Li, X.A., Das, S.K., Miften, M., Ten Haken, R.K. (2010). Radiation-associated liver injury. Int. J. Radiat. Oncol. Biol. Phys. 76**,** S94-S100.

Prayongrat, A., Umegaki, K., Van Der Schaaf, A., Koong, A.C., Lin, S.H., Whitaker, T., Mcnutt, T., Matsufuji, N., Graves, E., Mizuta, M., et al. (2018). Present developments in reaching an international consensus for a model-based approach to particle beam therapy. J. Radiat. Res. *59***,** I72-I76.

Prayongrat, A., Kobashi, K., Ito, Y.M., Katoh, N., Tamura, M., Dekura, Y., Toramatsu, C., Khorprasert, C., Amornwichet, N., Alisanant, P., Shirato, H., Shimizu, S. (2019). The normal tissue complication probability model-based approach considering uncertainties for the selective use of radiation modality in primary liver cancer patients. Radiother. Oncol. *135***,** 100-106.

Qi, W.X., Fu, S., Zhang, Q., Guo, X.M. (2015). Charged particle therapy versus photon therapy for patients with hepatocellular carcinoma: A systematic review and meta-analysis. Radiother. Oncol. *114*, 289-295.

Randall, K., Ten Haken, P.D. Individualizing radiotherapy using NTCP models: 17 years in Ann Arbor. [online]. USA: Association of Physicists in Medicine (AAPM). Available: http://www.aapm.org/meetings/amos2/pdf/29-7789-81574-66.pdf [Accessed Aug 30 2017].

Reed, G.B., Cox, A.J. (1966). The human liver after radiation injury. A form of veno-occlusive disease. Am. J. Pathol. *48*, 597-611.

Schaue, D., Kachikwu, E.L., Mcbride, W.H. (2012). Cytokines in radiobiological responses: A review. Radiat. Res. *178*, 505-523.

Schilstra, C., Meertens, H. (2001). Calculation of the uncertainty in complication probability for various dose-response models, applied to the parotid gland. Int. J. Radiat. Oncol. Biol. Phys. *50*, 147-158.

Schneider, U., Sumila, M., Robotka, J. (2011). Site-specific dose-response relationships for cancer induction from the combined Japanese a-bomb and Hodgkin cohorts for doses relevant to radiotherapy. Theor. Biol. Med. Model. *8*, 27.

Seong, J., Park, H.C., Han, K.H., Chon, C.Y. (2003). Clinical results and prognostic factors in radiotherapy for unresectable hepatocellular carcinoma: a retrospective study of 158 patients. Int. J. Radiat. Oncol. Biol. Phys. *55*, 329-336.

Seppenwoolde, Y., Lebesque, J.V., De Jaeger, K., Belderbos, J.S., Boersma, L.J., Schilstra, C., Henning, G.T., Hayman, J.A., Martel, M.K., Ten Haken, R.K. (2003). Comparing different NTCP models that predict the incidence of radiation pneumonitis. Int. J. Radiat. Oncol. Biol. Phys. *55*, 724-735.

Shirato, H., Le, Q.T., Kobashi, K., Prayongrat, A., Takao, S., Shimizu, S., Giaccia, A., Xing, L., Umegaki, K. (2018). Selection of external beam radiotherapy approaches for precise and accurate cancer treatment. J. Radiat. Res. *59*, I2-I10.

Sugahara, S., Nakayama, H., Fukuda, K., Mizumoto, M., Tokita, M., Abei, M., Shoda, J., Matsuzaki, Y., Thono, E., Tsuboi, K., et al. (2009). Proton-beam therapy for hepatocellular carcinoma associated with portal vein tumor thrombosis. Strahlenther. Onkol. *185*, 782-788.

Tai, A., Erickson, B., Li, X.A. (2009). Extrapolation of normal tissue complication probability for different fractionations in liver irradiation. Int. J. Radiat. Oncol. Biol. Phys. *74*, 283-289.

Tamura, M., Sakurai, H., Mizumoto, M., Kamizawa, S., Murayama, S., Yamashita, H., Takao, S., Suzuki, R., Shirato, H., Ito, Y. M. (2017). Lifetime attributable risk of radiation-induced secondary cancer from proton beam therapy compared with that of intensity-modulated x-ray therapy in randomly sampled pediatric cancer patients. J. Radiat. Res. *58*, 363-371.

Tanguturi, S.K., Wo, J.Y., Zhu, A.X., Dawson, L.A., Hong, T.S. (2014). Radiation therapy for liver tumors: Ready for inclusion in guidelines? Oncologist *19*, 868-879.

Tao, C., Yang, L.X. (2012). Improved radiotherapy for primary and secondary liver cancer: stereotactic body radiation therapy. Anticancer Res. *32*, 649-655.

Team, R.D.C. 2010. R: A language and environment for statistical computing. (Vienna, austria: R foundation for statistical computing).

Tome, W.A., Fenwick, J.D. (2004). Analysis of radiation-induced liver disease using the lyman NTCP model: in regard to Dawson et al. Int. J. Radiat. Oncol. Biol. Phys. *58*, 1318-1319.

Toramatsu, C., Katoh, N., Shimizu, S., Nihongi, H., Matsuura, T., Takao, S., Miyamoto, N., Suzuki, R., Sutherland, K., Kinoshita, R., et al. (2013). 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. *8*, 48.

Wedenberg, M. (2013). Assessing the uncertainty in QUANTEC's dose-response relation of lung and spinal cord with a bootstrap analysis. Int. J. Radiat. Oncol. Biol. Phys. *87*, 795-801.

Widder, J., Van Der Schaaf, A., Lambin, P., Marijnen, C.A., Pignol, J.P., Rasch, C.R., Slotman, B.J., Verheij, M., Langendijk, J.A. (2016). The quest for evidence for proton therapy: model-based approach and precision medicine. Int. J. Radiat. Oncol. Biol. Phys. *95*, 30-36.

Xu, Z.Y., Liang, S.X., Zhu, J., Zhu, X.D., Zhao, J.D., Lu, H.J., Yang, Y.L., Chen, L., Wang, A.Y., Fu, X.L., et al. (2006). Prediction of radiation-induced liver disease by Lyman normal-tissue complication probability model in three-dimensional conformal radiation therapy for primary liver carcinoma. Int. J. Radiat. Oncol. Biol. Phys. *65*, 189-195.