OncoRay – National Center for Radiation Research in Oncology
Direktorin: Frau Prof. Dr. Mechthild Krause

Towards patient selection for cranial proton beam therapy – Assessment of current patient-individual treatment decision strategies

D i s s e r t a t i o n s s c h r i f t

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)
vorgelegt
der Medizinischen Fakultät Carl Gustav Carus
der Technischen Universität Dresden

von

M. Sc. Almut Dutz

aus Lichtenstein/Sa.

Dresden 2020

1. Gutachter: Prof. Dr. Steffen Löck

2. Gutachter: Prof. Dr. Johannes Albertus Langendijk

Tag der mündlichen Prüfung: 10. August 2020

gez.: Prof. Dr. Edmund Koch
Vorsitzender der Promotionskommission

Abstract

Proton beam therapy shows dosimetric advantages in terms of sparing healthy tissue compared to conventional photon radiotherapy. Those patients who are supposed to experience the greatest reduction in side effects should preferably be treated with proton beam therapy. One option for this patient selection is the model-based approach. Its feasibility in patients with intracranial tumours is investigated in this thesis. First, normal tissue complication probability models for early and late side effects were developed and validated in external cohorts based on data of patients treated with proton beam therapy. Acute erythema as well as acute and late alopecia were associated with high-dose parameters of the skin. Late mild hearing loss was related to the mean dose of the ipsilateral cochlea. Second, neurocognitive function as a relevant side effect for brain tumour patients was investigated in detail using subjective and objective measures. It remained largely stable during recurrence-free follow-up until two years after proton beam therapy. Finally, potential toxicity differences were evaluated based on an individual proton and photon treatment plan comparison as well as on models predicting various side effects. Although proton beam therapy was able to achieve a high relative reduction of dose exposure in contralateral organs at risk, the associated reduction of side effect probabilities was less pronounced. Using a model-based selection procedure, the majority of the examined patients would have been eligible for proton beam therapy, mainly due to the predictions of a model on neurocognitive function.

Kurzzusammenfassung

Die Protonentherapie zeigt dosimetrische Vorteile gegenüber konventioneller Photonentherapie in der Schonung gesunden Gewebes. Patienten, welche eine hohe Reduktion von Nebenwirkungen erwarten können, sollten bevorzugt mit einer Protonentherapie behandelt werden. Die Patientenauswahl kann durch den modellbasierten Ansatz geschehen, dessen Realisierbarkeit bei Patienten mit intrakraniellen Tumoren in dieser Arbeit untersucht wird. Dazu wurden, basierend auf Daten von mit Protonentherapie behandelten Patienten, zuerst Modelle zur Vorhersage von Nebenwirkungswahrscheinlichkeiten von Früh- und Spätfolgen entwickelt und an externen Kohorten validiert. Akute Erytheme sowie akute und späte Alopezie waren mit Hochdosisparametern der Haut assoziiert. Späte leichte Hörminderung zeigte einen Zusammenhang mit der mittleren Dosis der ipsilateralen Cochlea. Des Weiteren wurde die neurokognitive Funktion als relevante Nebenwirkung bei Hirntumorpatienten eingehend mit subjektiven und objektiven Methoden untersucht. Sie blieb innerhalb der rezidivfreien Zeit bis zu zwei Jahre nach Protonentherapie weitgehend stabil. Abschließend wurden anhand eines individuellen Vergleichs von Protonen- und Photonentherapieplänen sowie verschiedener Modelle potentielle Toxizitätsunterschiede evaluiert. Obwohl die Protonentherapie eine hohe relative Reduktion der Dosisbelastung in kontralateralen Risikoorganen erzielte, war die begleitende Reduktion der Nebenwirkungswahrscheinlichkeit weniger deutlich ausgeprägt. Bei Verwendung eines modellbasierten Verfahrens wäre die Mehrheit der untersuchten Patienten für eine Protonentherapie ausgewählt worden, vor allem aufgrund der Vorhersagen eines Modells zur neurokognitiven Funktion.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3D-CRT	three-dimensional conformal radiotherapy
AUC	area under the receiver operating characteristic curve
CI	confidence interval
CNS	central nervous system
CRT	chemoradiotherapy
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
CTV	clinical target volume
CTx	chemotherapy
DNA	deoxyribonucleic acid
DS	double scattering
DVH	dose-volume histogram
Dx%	dose in x% of the volume of a structure
ECOG	Eastern Co-operative of Oncology Group
EORTC	European Organisation for Research and Treatment of Cancer
FACT	Functional Assessment of Cancer Therapy
GEE	generalised estimating equation
gEUD	generalised equivalent uniform dose
GTV	gross tumour volume
HGG	high-grade glioma
HU	Hounsfield units
ICRU	International Commission on Radiation Units and Measurements
IDH	isocitrate dehydrogenase
IGRT	image-guided radiotherapy
IMRT	intensity-modulated radiotherapy
LENT-SOMA	late effects of normal tissues – subjective, objective, management and analytic
LET	linear energy transfer
LGG	low-grade glioma
LKB model	Lyman-Kutcher-Burman model
MGH	Massachusetts General Hospital

MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
MR	magnetic resonance
MRI	magnetic resonance imaging
NTCP	normal tissue complication probability
OAR	organ at risk
PBS	pencil-beam scanning
PBT	proton beam therapy
PCV	procarbazine-lomustine-vincristine
PET	positron emission tomography
PRO	patient-reported outcome
PTV	planning target volume
QLQ	quality of life questionnaire
QLQ-BN20	brain tumour-specific quality of life questionnaire
QLQ-C30	quality of life core questionnaire
QoL	quality of life
QUANTEC	Quantitative Analyses of Normal Tissue Effects in the Clinic
RBE	radiobiological effectiveness
RCT	randomised controlled trial
ROC	receiver-operating characteristic
RT	radiotherapy
RTOG	Radiation Therapy Oncology Group
SD	standard deviation
SOBP	spread-out Bragg peak
TMZ	temozolomide
UPTD	University Proton Therapy Dresden
VMAT	volumetric modulated arc therapy
VxGy	volume of a structure receiving x Gy
WHO	World Health Organization
WMS-WL	Wechsler Memory Scale-III Word List
WPE	West German Proton Therapy Centre Essen
XRT	conventional photon radiotherapy

1 Introduction

In the 1950s, the first cancer patients have been treated with proton beam therapy (PBT) (Smith, 2006). A superiority to conventional photon radiotherapy (XRT) is mainly expected due to the physical properties of protons. As charged particles, they interact differently with matter than photons and therefore stop after a certain range in tissue. Consequently, the healthy tissue behind the tumour is less exposed and high dose conformity can be achieved. These dosimetric advantages may be beneficial in two different ways. First, dose could be escalated in radioresistant tumours, increasing tumour control while keeping potential side effects at a level comparable to non-escalated XRT regimes. Second, radiation-induced side effects and the risk of secondary tumour induction could be generally reduced due to the lower exposure of healthy tissue. The clinical implementation of PBT was mainly based on the dosimetric superiority of protons shown in comparative *in silico* planning studies (Goitein and Cox, 2008; Suit et al., 2008). Although the dosimetric advantages of PBT over XRT are often apparent, the respective dose reduction does not always lead to a substantial reduction in side effects (Durante et al., 2017). Thus, the question is raised whether it is indeed appropriate to justify the use of PBT solely based on planning studies and the ALARA principle (as low as reasonably achievable), as previously done when introducing new radiotherapy (RT) techniques such as intensity-modulated radiotherapy (IMRT) (Jacobs et al., 2017). This question is accompanied by concerns about the considerably higher technical, personnel and therefore financial effort required to operate a PBT facility compared to conventional XRT (Lievens and Pijls-Johannesma, 2013; Bekelman and Hahn, 2014).

Despite high capital costs, which constitute an important limiting factor for the installation of a PBT facility (Lievens and den Bogaert, 2005), the number of PBT centres has increased rapidly in recent decades, especially since the 2000s (figure 1.1). Although many patients have already been treated with PBT, the published clinical evidence for its superiority over XRT is still limited (Grau, 2013). One reason for this lack of evidence could be that university-based facilities tend to focus on research in physics but less on patient care, while hospital-based facilities operate more on a commercial basis and have less academic focus (Grau, 2013). Commonly, a proof for the clinical superiority of PBT in terms of prospective randomised controlled trials (RCTs) is demanded (Zips and Baumann, 2013). However, it is controversial whether RCTs are suited to reveal potential differences between PBT and XRT (Glimelius and Montelius, 2007; Macbeth and Williams, 2008). In contrast to RCTs comparing systemic therapies, there are more degrees of freedom when comparing PBT with a conventional irradiation technique. There exists no *standard application* of RT, neither XRT (Fairchild et al., 2014; Das et al., 2017) nor PBT (Taylor et al., 2016). The heterogeneity between centres in terms of treatment planning systems, quality assurance routines, training skills, image guidance, treatment adaptation, immobilisation strategies

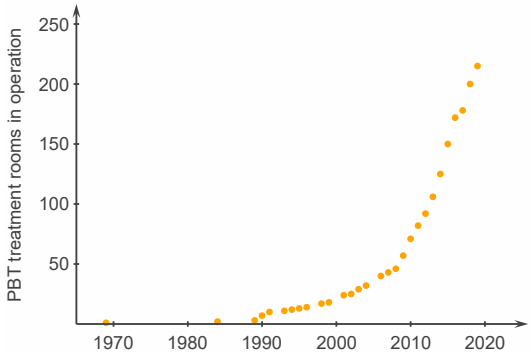


Figure 1.1: Number of proton beam therapy treatment rooms in clinical operation, data as of September 2019 (Particle Therapy Co-Operative Group, 2019).

etc. may be so pronounced that it can be difficult to transfer results from RCTs into clinical routine (Widder et al., 2016; Langendijk et al., 2018).

When comparing PBT and XRT in terms of reduced side effects, some additional obstacles may occur. Late radiation-induced side effects as relevant endpoints of such clinical trials may manifest many years following RT. Thus, results from large long-term RCTs may be obsolete as RT (and PBT in particular) is still a rapidly evolving technology (Langendijk et al., 2013). Usually, all eligible patients are enrolled in an RCT, regardless of their presumed extent of side effect reduction under PBTs. This requires the inclusion of a large number of patients to observe any significant difference between both treatment modalities. The number of included patients may be reduced by a proper preselection of eligible patients. Due to this preselection process, the estimated difference in complication probability between both treatment groups is enhanced, so that an adequate balance between risks and benefits for patients in both groups (equipoise) is supposed to be lacking (Bentzen, 2008). Consequently, RCTs may be considered unethical (Goitein and Cox, 2008; Suit et al., 2008).

As an alternative possibility for PBT patient selection, the model-based approach has been suggested and introduced into clinical practice in the Netherlands (Langendijk et al., 2013). In this approach, patients with non-standard indications are selected for PBT who could benefit most in terms of reduced side effects, see figure 1.2. In a first step, valid normal tissue complication probability (NTCP) models predicting certain radiation-induced side effects are developed and selected since they provide the basis for the whole approach. During the next step, two treatment plans, one using PBT and the other using state-of-the-art XRT, are created for each patient. Using the selected NTCP models and the two treatment plans, the complication risk of the relevant side effect can be estimated for each modality. The difference $\Delta NTCP$ determines the clinical benefit that can potentially be achieved if the patient is treated with protons instead of photons. If this estimate exceeds a predefined threshold, e.g. 10 percentage points, the patient is treated with PBT, otherwise with XRT. Besides patient selection, evaluation is an integral part of

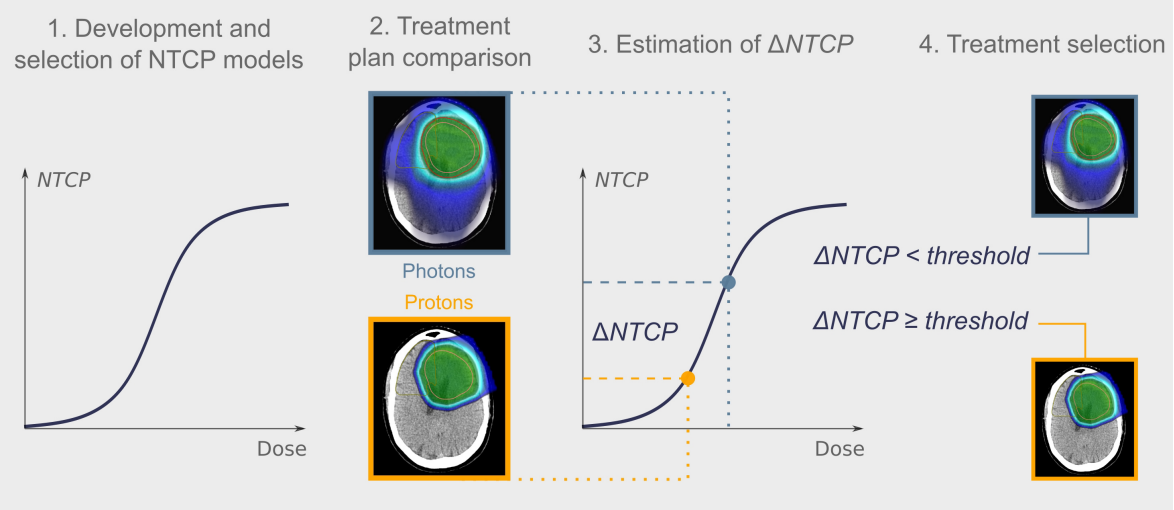


Figure 1.2: Schematic overview of the model-based patient selection for proton beam therapy. NTCP, normal tissue complication probability. Adapted from Langendijk et al. (2013).

the model-based approach. Only by systematically selecting suitable patients for PBT, the PBT cohort can be enriched to generate evidence for PBT more efficiently (Widder et al., 2016).

This approach has mainly been implemented for patients with head and neck cancer, as a variety of validated NTCP models for several side effects have been developed for this tumour entity, such as xerostomia, sticky saliva or tube feeding (Beetz et al., 2012a; Beetz et al., 2012b; Beetz et al., 2012c; Wopken et al., 2014). For other tumour entities, fewer validated NTCP models for radiation-induced side effects are available. The prediction of NTCP models may vary depending on the patient population (Peeters et al., 2006) or the applied RT techniques (Beetz et al., 2012b; Christianen et al., 2012; Troeller et al., 2015; Pedersen et al., 2020). Almost all models have been developed using XRT data, although it is not entirely certain whether all of these models can also be applied to PBT.

Cranial RT may be associated with mild to severe side effects such as alopecia, hearing impairment, neurocognitive and endocrine dysfunctions, cranial nerve failure or brain necrosis (Shih et al., 2009; Landier, 2016). These side effects can affect patients' quality of life (QoL) adversely. Patients with benign tumours or low-grade gliomas (LGGs) often have prolonged survival and young age at diagnosis (Buckner et al., 2016). These long-term survivors may be affected by late side effects. Patients with high-grade gliomas (HGGs) have a short life expectancy (Stupp et al., 2009) and thus, an important goal of treatment is the preservation of QoL as long as possible by reducing early side effects. Comparative dosimetric studies in brain tumour patients showed dosimetric advantages of PBT over XRT also in these entities (Harrabi et al., 2016; Adeberg et al., 2018). A significant dose reduction, especially in organs at risk (OARs) associated with neurocognitive function and QoL, promises potential for reducing late side effects. So far, treatment decisions for patients with brain tumours have been based mainly on dosimetric comparisons.

Alternatively, the selection of these patients for PBT patients may be realised using the model-based approach. For XRT, NTCP models for different endpoints have been developed, such as neurocognitive function, radiation-induced brain necrosis, tinnitus, and ocular toxicity (Burman et al., 1991; Bender, 2012; Gondi et al., 2012; Batth et al., 2013; Lee et al., 2015). However, studies based on data from PBT or combined PBT and XRT are rare (De Marzi et al., 2015; Palma et al., 2020).

This thesis investigates whether the model-based approach can be applied to patients with intracranial tumours. In a multi-centre approach, data from a total of 305 patients treated with cranial PBT at one of three different PBT facilities were evaluated. NTCP models for common early and late physician-rated side effects were developed and validated. The preservation of neurocognitive function is of paramount importance for brain tumour patients and was additionally investigated in more detail using subjective patient-reported as well as objective measures. Finally, the potential of dose and NTCP reduction in comparison to XRT was estimated based on an *in silico* treatment plan comparison.

2 Theoretical background

2.1 Treatment strategies for tumours in the brain and skull base

The most common types of adult intracranial tumours and their management are discussed in this section with focus on RT, particularly PBT. All primary brain tumours originate from cells within the central nervous system (CNS) and account for 2 % of all cancer types. About two-thirds of primary brain tumours are non-malignant, see figure 2.1. The World Health Organization (WHO) classified tumours of the central nervous system into different grades based on histology and molecular parameters (Louis et al., 2016). Low-grade meningiomas are the most common non-malignant brain tumours with a 5-year overall survival of 87 %. Almost half (47 %) of all malignant brain tumours are glioblastomas. They are considered extremely difficult to treat as their 5-year overall survival is only 5.5 % (Ostrom et al., 2017).

Besides clinical and genetic parameters, differential diagnosis is primarily based on multimodal magnetic resonance imaging (MRI) including imaging with contrast agent, diffusion-weighted or diffusion tensor imaging, magnetic resonance (MR) perfusion and spectroscopy. Diagnostic images contribute to a better definition of the tumour or surgical cavity, surrounding oedema and healthy tissues. Subsequently, MR images are used in the phase of delineation of treatment volumes and OARs for RT treatment planning.

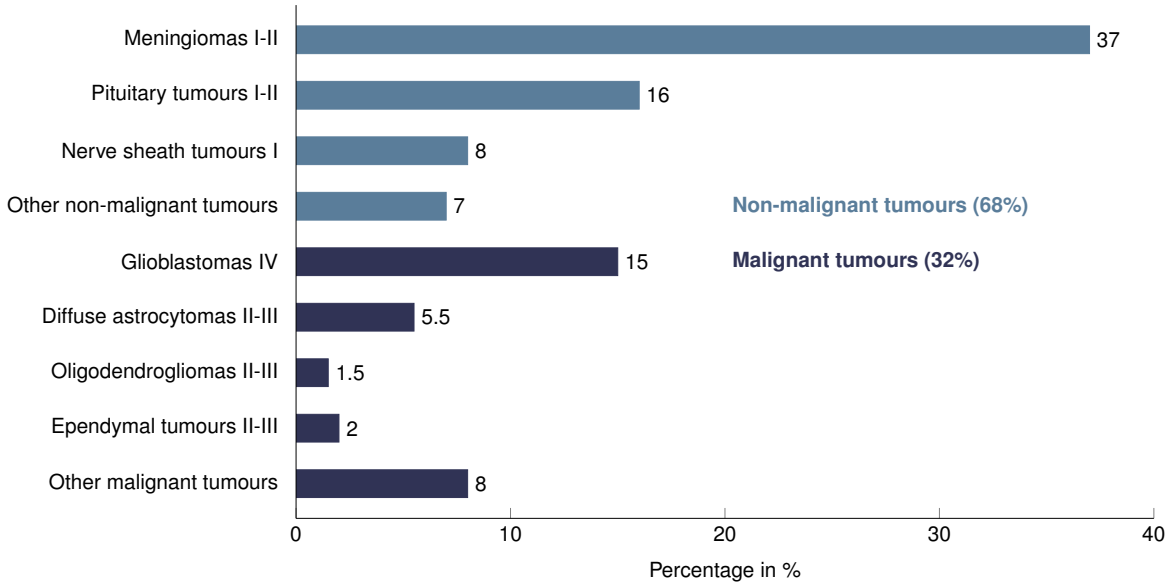


Figure 2.1: Yearly incidence rates of primary brain tumours. Adapted from Ostrom et al. (2017). WHO grades are given in Roman numerals.

2.1.1 Gliomas

Gliomas arise from glial or precursor cells and comprise glioblastomas, astrocytomas, and oligodendrogliomas. Primary treatment of most gliomas is surgery aiming at maximum resection safety, as the extent of resection determines survival and functional status in LGG (van den Bent et al., 2017b) and HGG (Chaichana et al., 2014; Noorbakhsh et al., 2014). Postoperative treatment depends on tumour classification and risk stratification of the patient. It includes either RT or chemotherapy (CTx) or a combination of both. Unresectable tumours are generally treated with chemoradiotherapy (CRT).

Diffuse low-grade gliomas (WHO grade II) The 5-year survival rate of patients with diffuse oligodendrogliomas grade II and diffuse astrocytomas grade II are 81 % and 50 %, respectively (Lapointe et al., 2018). For low-risk patients, the European Organisation for Research and Treatment of Cancer (EORTC) 22845 trial recommends a postoperative watch-and-wait with regular MRI follow-up (van den Bent et al., 2005; Weller et al., 2017). For isocitrate dehydrogenase (IDH)-mutant high-risk patients, standard of care is adjuvant RT with doses of 50 – 54 Gy followed by chemotherapy with procarbazine-lomustine-vincristine (PCV) or temozolomide (TMZ) (Karim et al., 1996; Shaw et al., 2002; Buckner et al., 2016; Weller et al., 2017).

Diffuse high-grade gliomas (WHO grade III) The 5-year survival rate of patients with anaplastic oligodendrogliomas grade III and anaplastic astrocytomas grade III are 57 % and 30 %, respectively (Lapointe et al., 2018). Patients with high-grade gliomas are treated with postoperative combination treatment including radiation with doses of 60 Gy and adjuvant chemotherapy with PCV. For specific patient groups, the combination of RT and adjuvant chemotherapy with TMZ improves overall survival (van den Bent et al., 2017a).

Glioblastomas (WHO grade IV) Glioblastomas are the most aggressive brain tumours. On average, only 5.5 % of glioblastoma patients are still alive 5 years after diagnosis (Ostrom et al., 2017). Standard of care of patients with glioblastomas and Karnofsky performance status ≥ 60 is adjuvant CRT including RT with doses of 60 Gy and concomitant and adjuvant TMZ (Stupp et al., 2005). For elderly patients (age ≥ 70 years), a hypofractionated RT regimen with a total dose of 40.05 Gy in 15 daily fractions with or without concomitant and adjuvant TMZ reduces treatment time (Perry et al., 2017).

2.1.2 Meningiomas

Meningiomas are commonly slow-growing and well-circumscribed tumours that originate from the meninges (Whittle et al., 2004). Most meningiomas are benign (90 %, WHO grade I), but they also occur in atypical (5 – 7 %, WHO grade II) and anaplastic (1 – 3 %, WHO grade III) forms (Whittle et al., 2004). Meningiomas can reach considerable sizes and displace the brain of the

contralateral hemisphere. Treatment varies depending on the WHO grade, site and size of the lesion, symptoms and age of the patient (Wilson, 1994; Black et al., 2001). Primary treatment of operable patients is surgical resection (Whittle et al., 2004). Curative RT is indicated in case of inoperability or if surgery is refused by the patient. Incompletely resected or inoperable atypical or anaplastic meningiomas are indications for RT with doses up to 60 Gy (Hug et al., 2000). Symptomatic patients or progressive meningiomas grade I may benefit from normo-fractionated RT with doses of 50 – 54 Gy (Rogers et al., 2015). Stereotactic radiosurgery or fractionated stereotactic RT is a primary treatment option for small meningiomas grade I (Rogers et al., 2015). As RT can achieve a similar outcome in terms of local tumour control as surgery, RT may be considered as primary treatment option for small tumours (Pollock et al., 2003).

2.1.3 Pituitary adenomas

Pituitary adenomas are classified according to their size on MRI (Raverot et al., 2014). About one-third of pituitary adenomas are clinically nonfunctioning adenomas. Functioning adenomas secrete excess hormones, such as prolactin (prolactinoma), growth hormone or adrenocorticotropic hormone (Molitch, 2017). The initial treatment for most pituitary adenomas except for prolactinomas is transsphenoidal surgical resection of the tumour. The first treatment option for prolactinomas are dopamine agonists (Casanueva et al., 2006). In the cases of inoperability, tumour recurrence or progression after surgery or in patients with functioning pituitary adenomas with increased hormone levels after medical therapy and radical resection, RT is indicated (Loeffler and Shih, 2011). Single fraction stereotactic radiosurgery with doses of 14 – 18 Gy is suitable for microadenomas or small macroadenomas that are not in the immediate vicinity of the optic nerves or chiasm. Pituitary adenomas are irradiated with normo-fractionated doses between 45.0 – 54.0 Gy (Loeffler and Shih, 2011).

2.1.4 Tumours of the skull base

This term is used to describe both lesions of the skull base that can invade the cranial cavity, neck or face and lesions of the brain, face and neck that are adjacent to the skull base (Mazzoni and Krengli, 2016; Kunimatsu and Kunimatsu, 2017). Since the skull base consists of a multitude of different tissues such as meninges, bone, cartilage, soft tissue, nerves and nerve sheaths, benign and malignant skull base tumours have extremely heterogeneous histologies. Surgery and RT are the most relevant treatment options for skull base tumours. As these tumours are often located in the proximity of vital structures such as the brain stem, optic pathways, inner and middle ear, cranial nerves and major vessels, complete resection is limited in many cases. Post-operative RT is applied either as stereotactic RT for small and simple-shaped lesions or with highly conformal RT techniques for larger lesion and complex shapes (Fossati et al., 2016).

2.1.5 Role of proton beam therapy

No RCTs are available confirming any clinical benefit of PBT over XRT in CNS tumours (Combs, 2017). Use of PBT in brain tumours is mainly justified with a potential reduction of side effects, i.e. neurocognitive sequelae or secondary malignancies (Combs, 2017). Tumours located in the immediate vicinity of OARs, such as brain stem, chiasm or optic nerves, may benefit from PBT. Due to the sharp dose gradient close to the tumour, high doses can be applied to the tumour while sparing OARs. In patients with LGG treated with PBT, stable neurocognitive function and QoL was observed (Sherman et al., 2016). A multicentre study on patients with LGG reported acute side effects of grade 1 and 2 following PBT: 81 % alopecia, 78 % dermatitis, 47 % fatigue and 40 % headache. No acute side effects grade $\geq$3 were observed (Wilkinson et al., 2016).

Moreover, PBT offers the option of dose escalation that may lead to increased tumour control for HGGs (Fitzek et al., 2001), atypical meningiomas (McDonald et al., 2015) and for chordoma and chondrosarcoma of the skull base (Palm et al., 2019). Since radioresistant chordomas and chondrosarcomas have to be treated to high radiation doses ($>$ 70 Gy) but are adjacent to critical organs, they were among the first entities to be irradiated with protons (Mazzoni and Krengli, 2016). Due to the more favourable outcome of PBT (5-year locoregional control of 46 – 81 %) compared to XRT (15 – 65 %), PBT is considered the gold standard for chordomas (Fossati et al., 2016; Combs, 2017).

In case of re-irradiation of tumour recurrences, the applied total dose to the tumour may have to be reduced as the normal tissue may have been damaged during the primary course of RT. To reduce the dose exposure of the pre-irradiated normal tissue during re-irradiation, PBT could be a treatment option (Glimelius et al., 2009).

Because available evidence is still limited, especially concerning to long-term side effects, there are a few ongoing trials comparing proton and photon RT of different techniques. These include preference-based non-randomised studies (NCT02824731, 2016) and RCTs for grade II/III glioma (NCT03180502, 2017) and glioblastoma (NCT01854554, 2013; NCT02179086, 2014).

2.2 Radiotherapy with photons and protons

The effect of RT treatment is based on energy transfer to tumour cells to prevent further tumour growth and subsequently eliminate the malignant cells. The mean absorbed energy per unit mass is defined as energy dose D. Both photons and charged particles are used to deposit dose in the tumour. Because of fundamental differences in their physical properties, the resulting dose distributions vary mainly in their depth-dose distributions. This section summarises the basic biological effects of radiation as well as physical principles of RT with charged and uncharged particles and their impact on treatment planning for both radiation types used in this thesis: photons and protons.

2.2.1 Biological effect of radiation

The critical target of radiation-induced tumour cell death is deoxyribonucleic acid (DNA) (Wouters and Begg, 2009). DNA damage may inhibit cell proliferation and thus tumour growth and metastatic spread. Radiation can directly ionise molecules within the DNA but more often the DNA is indirectly affected. Radiation may create chemical radicals, e.g. by ionising cellular water, that in turn damages the molecular structure. The most effective damages to DNA are complex or clustered double-strand breaks as the probability of repair is low. Particles with high linear energy transfer (LET) will more likely cause complex double-strand breaks due to multiple interactions with the same structure alongside a single particle trajectory. For low-LET radiation, double-strand breaks are more likely induced by single sublethal damages of several independent particle trajectories. Accumulation of multiple sublethal damages can finally cause lethal damage (Krieger, 2012).

Tumour cells are less efficient in repair of DNA damage compared to late-responding normal tissue cells with slow cell division. Thus, the total dose is generally applied in fractions to allow healthy cells to recover during the time between daily fractions. Standard fractionation for curative treatment is 1.8 – 2.0 Gy per fraction, five fractions a week (Baumann and Grégoire, 2009). The biological effect of different fractionation schemes can be calculated with the linear-quadratic model, that characterises the survival fraction S/S_0 of an original cell population S_0 irradiated with dose D by

$$S/S_0(D) = \mathrm{e}^{-(\alpha D + \beta D^2)}, \tag{2.1}$$

where α and β are tissue-specific parameters describing the mechanisms of cell damage (Joiner, 2009). The ratio α/β characterises the steepness and bend of the survival curve. The α/β ratio is a widely used parameter to characterise the radiosensitivity and repair capacity of tumour and normal tissues: the greater α/β the lower its repair capacity. Since the α/β ratio for tumours is considerably higher than for late responding normal tissues, healthy tissue can be spared from lethal damage. The biological effect of different fractionation schemes is determined by converting the total dose into a biologically equivalent dose delivered in 2 Gy fractions. This equivalent dose EQD_2 is calculated by

$$EQD_2 = D \cdot \frac{\alpha/\beta + d}{\alpha/\beta + 2\,\mathrm{Gy}} \tag{2.2}$$

where d is the dose per fraction in Gy and D is the total dose of the alternative fractionation scheme (Joiner and Bentzen, 2009).

To compare the biological effects of different types of radiation, the radiobiological effectiveness (RBE) was introduced. RBE is the ratio of the absorbed dose of a reference radiation type D_{ref}

and the absorbed dose of the investigated radiation type D_{test}, which is required for the same biological effect under otherwise identical experimental conditions (Krieger, 2012)

$$RBE = \left.\frac{D_{\text{ref}}}{D_{\text{test}}}\right|_{\text{same biological effect}} . \tag{2.3}$$

Photon radiation is often set as a reference. RBE does not only depend on the radiation type and the specific biological effect, but also on LET and dose. For proton beams in clinical applications, RBE is considered constant at 1.1. However, with increasing LET along the proton path, RBE increased up to 1.7 at the distal dose fall-off in preclinical studies (Paganetti, 2014). Although these RBE variations are reported (Paganetti and van Luijk, 2013; Lühr et al., 2018; Willers et al., 2018), to date, a constant RBE has been accepted and used in clinical practice (Paganetti et al., 2002). Hence, in PBT, the physical dose with its unit Gy is weighted by the RBE factor 1.1 and the unit of this biological dose is renamed to Gy(RBE).

2.2.2 Basic physical principles of radiotherapy

Energy transfer and deposition in exposed tissues is caused by interactions of radiation with matter. These interactions differ depending on whether the radiation particles are charged or not. Charged particles such as protons ionise directly, i.e. the radiation field interacts directly with biomolecules. Uncharged particles, however, ionise indirectly, i.e. effects on biomolecules are caused by radiation-induced free water radicals (Krieger, 2012).

Photon radiotherapy An electron linear accelerator is used to produce high-energy photon beams with nominal beam energies between 4 MeV and 18 MeV. In historical analogy to diagnostic X-ray tubes, nominal beam energies are usually expressed in terms of acceleration voltages in the unit MV. To produce high-energy photons, electrons are accelerated in a waveguide and deflected onto a tungsten target. There, the kinetic energy of the electrons is converted into photon energy (bremsstrahlung). Photons in the given energy range interact with matter mainly by incoherent scattering. In such a process, a photon collides with a shell electron of an atom in the tissue, thereby transferring a part of the photon energy onto the scattered electron. The collision causes the electron to recoil and a new low-energy photon to be emitted into a different direction. This recoil electron may ionise many other atoms in the surrounding matter. Further interactions of photons with matter are coherent scattering, photoelectric effect, pair and triplet production and photodisintegration (Krieger, 2012). However, the probabilities for such interactions are very low for therapeutic photon energies and the atomic composition of human tissue. As photons pass through matter, the photon fluence is continuously attenuated according to the Beer-Lambert law. Thus, a photon beam has no finite range. These properties are reflected in the depth dose curve, see figure 2.2. In the built-up region, the dose increases to the dose maximum at a depth of a few centimetres depending on the initial photon energy and then decreases

exponentially (Krieger, 2012). Due to these physical properties, it is impossible to deliver high doses to deep-seated tumours while simultaneously sparing normal tissues with one single beam direction. Thus, the clinical dose is applied by multiple beams from different directions, whereby the dose contributions of every single beam superimpose in the target.

Proton beam therapy Protons for clinical applications are accelerated to therapeutical energies up to 250 MeV using circular accelerators such as cyclotrons (Krieger, 2012). In contrast to photons, charged particles usually lose their kinetic energy in many individual steps as they pass through matter and are slowed down continuously until they stop at a finite range depending on their initial energy. Energy losses are mainly caused by multiple interactions with the electron shells (ionisation and excitation) and by generating bremsstrahlung in the electric field of the atomic nuclei. For therapeutic proton energies and for the atomic composition of human tissue, stopping of protons is dominated by interactions with electrons.

The loss of energy per unit path length is defined as the stopping power of a material. Most of the initial proton energy is transferred at the end of its trajectory. This region of high dose is called Bragg peak (Krieger, 2012). Its location is determined by the initial proton energy, see figure 2.2. If beams with different initial proton energies and intensities are superimposed, the high dose region can be extended to cover an entire tumour volume from a single beam direction. This area is called spread-out Bragg peak (SOBP). As protons stop entirely after travelling a certain distance, the normal tissue behind the Bragg peak can be almost completely spared from dose. Thus, tumour coverage and sparing of healthy tissue can be achieved with few beams.

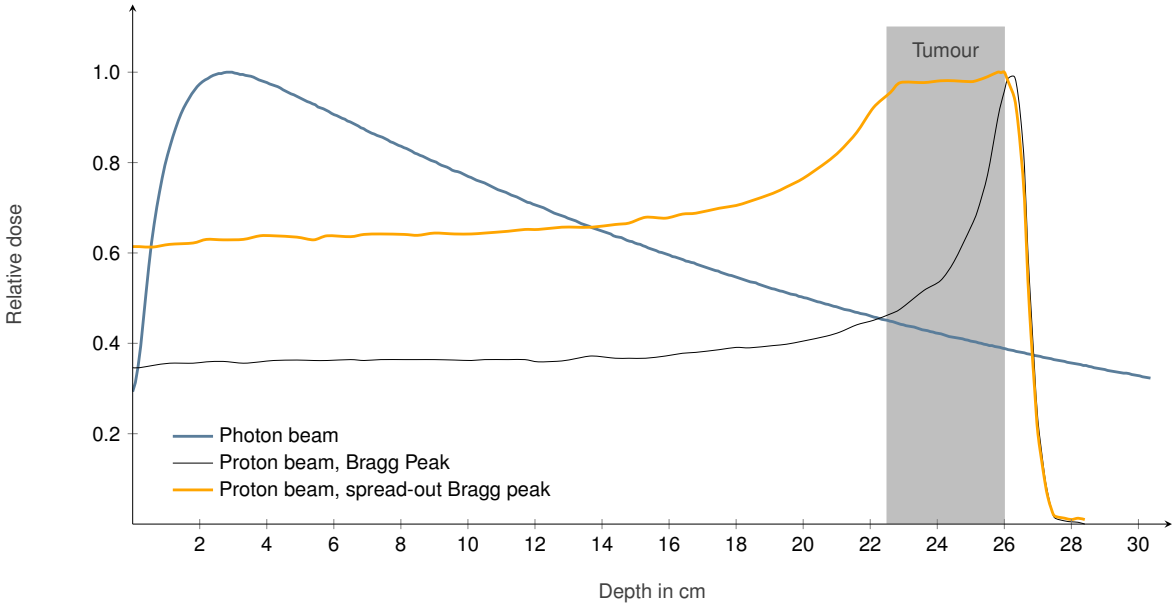


Figure 2.2: Depth dose curves of photon and proton beams in water. A monoenergetic proton beam and a spread-out Bragg peak (superimposition of monoenergetic proton beams of different energies) are shown.

2.2.3 Field formation in radiotherapy

The techniques of radiation therapy have been continuously developed to achieve a more conformal dose distribution to spare healthy tissue. This process was mainly driven by the introduction of three-dimensional image modalities and multi-leaf collimators as well as the development of inverse planning techniques.

Photon radiotherapy XRT started with single conformal fields in simple anterior-posterior beam configurations. With the introduction of computed tomography (CT) and sufficient dose computation capacities, three-dimensional conformal radiotherapy (3D-CRT) was developed. In this technique, single fields with a homogeneous beam profile are applied from several determined beam directions. This homogeneous beam profile is created by a flattening filter inserted in the beam directly behind the tungsten target. Each field is laterally collimated to the target volume using a multi-leaf collimator.

Nowadays, 3D-CRT is more and more replaced by inverse planning techniques with inhomogeneous fluence profiles. In these techniques, a single field is divided into multiple segments. For each field segment, the dose rate as well as the dose output is modulated. The segment is shaped using a multi-leaf collimator. The radiation technique is called IMRT if several fixed beam directions are used or volumetric modulated arc therapy (VMAT) if the radiation is delivered while the modulated beam rotates around the patient, see figure 2.3. Inverse techniques are seen as state-of-the-art XRT as they offer high conformity while sparing normal tissues.

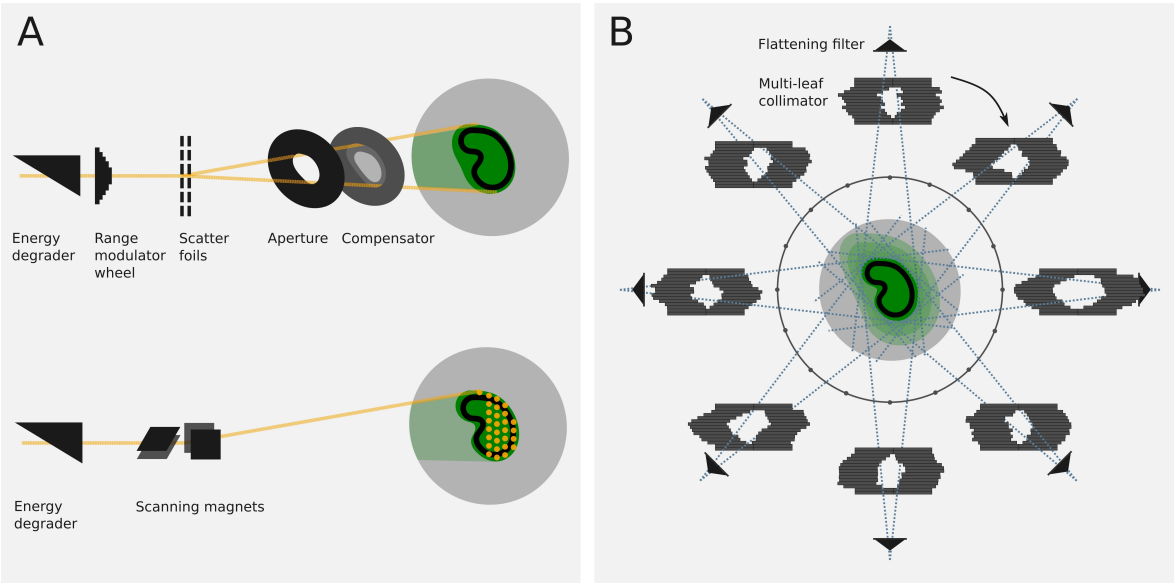


Figure 2.3: Radiotherapy techniques. **A** Proton beam therapy in double scattering (top) and pencil-beam scanning technique (bottom). **B** In photon volumetric modulated arc therapy (VMAT), radiation is delivered while the modulated beam rotates around the patient. Adapted from Jakobi (2015).

Proton beam therapy Two different PBT delivery methods were applied in this work: passive double scattering (DS) and active pencil-beam scanning (PBS), see figure 2.3. In DS, an SOBP is generated by a set of range modulator wheels or ridge filters that produce a spectrum of proton energies. The SOBP covers the target volume in beam direction. Since the beam generated in the cyclotron is very narrow, it is broadened by two scattering materials to cover the target laterally with a uniform dose profile. A field-specific collimator (aperture) conforms the dose laterally to the target volume and a milled range compensator aligns the dose to the distal edge of the target. With this technique, the dose cannot be adjusted to the proximal edge of the target.

In PBS techniques, narrow monoenergetic proton beams are deflected by sweeper magnets. Each monoenergetic beam scans the target volume along a two-dimensional grid in a determined energy layer. To cover the entire target volume, successive energy layers are scanned in beam direction. The penetration depth of the Bragg peak is adjusted by varying the beam energy using an energy degrader. Since the proton minimum energy is fixed due to technical constraints, targets directly below the body surface cannot be covered. For superficial tumours, a uniform material is inserted into the beam to reduce its energy. According to its function, this material (e.g. plastic) is called range shifter (Shen et al., 2015).

2.2.4 Target definition and delineation of organs at risk

Imaging modalities CT scans are performed in preparation of RT treatment planning. They contain information on tissue electron density. The anatomy is visualised by grey values in Hounsfield units (HU). The HU scale represents the photon attenuation coefficient of tissue in relation to that of water. The tumour as well as several OARs are delineated based on the planning CT scan. For optimal soft tissue differentiation, information drawn from MRI scans are considered. Commonly used sequences for the delineation of brain tumours are T1-weighted, T2-weighted and fluid-attenuated inversion recovery (FLAIR) with and without contrast agent (gadolinium). A functional imaging technique that visualises metabolic processes depending on the applied biologically active tracer molecule is positron emission tomography (PET). For brain tumour patients, the amino acid PET tracer ^{11}C-methionine is frequently used in some centres (Glaudemans et al., 2013).

Target definition The delineation of the target volume based on several diagnostic modalities is an essential step in the RT planning process. All macroscopic tumour volumes (primary tumour, metastatic regional nodes or distant metastases) are delineated in the gross tumour volume (GTV). In addition to the macroscopic tumour volume, the clinical target volume (CTV) also includes volumes in which proliferating tumour cells (microscopic tumour) are suspected (ICRU, 2010).

In general, the CTV underlies intra- and interfractional spatial displacements due to e.g. organ movement, daily patient setup variations, dose calculation approximations and uncertainties in

beam delivery. To ensure that the prescribed dose is still delivered to the CTV, the International Commission on Radiation Units and Measurements (ICRU) introduced the planning target volume (PTV) concept for XRT (ICRU, 1999). The PTV is created by applying safety margins to the CTV. Margin sizes vary depending on tumour location with respect to moving organs, applied radiation technique, image-guided radiotherapy (IGRT) regimes and patient immobilisation devices.

For PBT, however, there is a further uncertainty in beam direction. This is due to the PBT dose calculation algorithm which uses the proton stopping power of the traversed materials. For this sake, photon derived HU values are converted into relative stopping power values in the planning CT scan. This is achieved by conversion schemes having uncertainties (Paganetti, 2012). The uncertainty is different in lateral direction to that in beam direction and depends on the location of the tumour. Thus, the PTV concept of XRT cannot easily be employed. Instead, non-uniform margins are applied. Treatment planning with DS technique assumes an uncertainty in the proton beam range of a specific percentage plus an additional absolute margin (Paganetti, 2012). For PBS, these uncertainties are included in robust optimisation and evaluation (Liu et al., 2012).

Delineation of organs at risk Delineation of OARs is essential for optimisation and assessment of a treatment plan. For patients with brain tumours, the following OARs should be considered: brain stem, chiasm, optic nerves, lacrimal glands, eye lenses, cochleae or inner ears, brain excluding the target volume, and depending on the tumour location spinal cord and salivary glands (Wright et al., 2019). For treatment planning and assessment as well as consistency in multi-centre clinical trials, OARs should be contoured according to standardised guidelines on CT and MR images. For the sake of NTCP modelling or retrospective studies, it may be necessary to delineate further OARs.

2.2.5 Treatment plan assessment

In general, an RT treatment plan of an individual patient is assessed with regard to a balance of risks and benefits based on clinical experience: Benefits in terms of sufficient tumour coverage and risks in terms of acceptable doses to normal tissues. For treatments with curative intent, underdosage of the tumour may be more serious than normal tissue complications as local tumour recurrences are difficult to manage (Marks et al., 2010). The dose distribution is visualised with isodoses to evaluate target coverage and dose conformity as well as to detect undesired dose maxima, see figure 2.4.

The commonly used tool to evaluate dosimetric parameters of target volumes and OARs is the dose-volume histogram (DVH). Figure 2.5 shows an exemplary DVH of an OAR for two different treatment plans, one VMAT and one PBS plan. The ICRU proposed a standardised assessment of treatment plans (ICRU, 1999; ICRU, 2007; ICRU, 2010). Common criteria are the volume of a structure receiving x Gy (*VxGy*) and the dose in x% of the volume of a structure (*Dx%*). To achieve sufficient tumour coverage and avoid high doses exceeding the prescribed dose, mean,

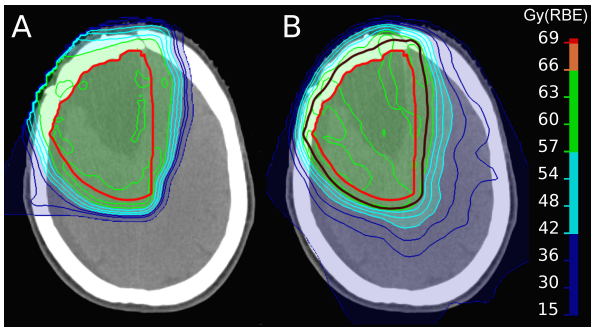


Figure 2.4: Dose distributions for a patient with oligodendroglioma grade III. The clinical and planning target volumes are delineated in red and brown, respectively. **A** Proton treatment plan created with double-scattering technique. **B** Photon treatment plan created with VMAT technique.

minimum and maximum dose values in the target should be close to the prescribed dose. Therefore, the near-minimum and near-maximum parameters $D98\%$ and $D2\%$ of the target volumes are usually evaluated. Additionally, the parts of the target volumes receiving 95 % or 107 % of the prescribed dose ($V95\%$ and $V107\%$) are assessed. To evaluate how well the dose distribution conforms to the size and shape of the target volume, conformity indices are calculated. A widely used conformity index CI was developed by Paddick (2000):

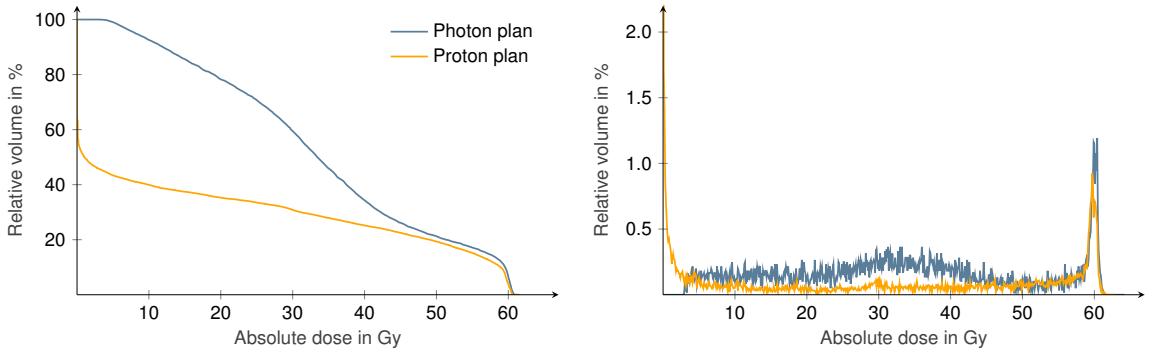
$$CI = \frac{TV_{\text{isodose}}^2}{TV \cdot V_{\text{isodose}}}, \tag{2.4}$$

where TV_{isodose} represents the part of the absolute target volume (CTV or PTV) receiving the considered dose ($V95\%$ or $V98\%$ of the target volume). TV describes the absolute target volume and V_{isodose} the absolute volume of the external structure (body) exposed to the considered dose. $CI = 1$ indicates 100 % conformity, while $CI = 0$ means total failure.

The homogeneity index HI characterises the uniformity of the dose distribution in the target: the lower the score, the higher the uniformity of the plan. It is defined by the ICRU (2010) as

$$HI = \frac{D2\% - D98\%}{D50\%}, \tag{2.5}$$

where $D2\%$, $D98\%$ and $D50\%$ refer to the investigated target volume (CTV or PTV). To evaluate the risk of radiotoxic normal tissue doses, DVH parameters are compared to clinical population-based constraints for different OARs. Such constraints have been summarised in organ-specific publications by Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) (Marks et al., 2010). Although a number of NTCP models for certain complications have been developed, the calculation of numerical complication probability for individual patients is rarely done in clinical routine (Bentzen et al., 2010).



(A) Cumulative dose volume histogram.

(B) Differential dose volume histogram.

Figure 2.5: (A) Cumulative and **(B)** the corresponding differential DVH of the ipsilateral frontal lobe for a patient with astrocytoma treated to 60 Gy(RBE). Proton treatment was planned using a DS technique and the comparable photon treatment was planned using a VMAT technique.

2.3 Patient outcome

As many types of cancer are life-threatening, tumour control has a high priority in cancer treatment. However, radical treatment is often limited as surrounding healthy organs will lose their functions. Already in the early years of RT, Holthusen (1936) described the probability of achieving tumour control and developing normal tissue damage following RT as a function of radiation dose, see figure 2.6. The difference between doses required for tumour control and tissue tolerance doses is referred to as the therapeutic window (Joiner et al., 2009). Efforts have been made to widen the therapeutic window by increasing tumour control and decreasing the risk of side effects by introducing new technologies or biological modulation (Combs, 2017). The different measures of tumour control, as well as normal tissue complications, are explained in this section.

2.3.1 Scoring of side effects

Radiation dose does not only affect the tumour but also damages normal tissue. Over time, many centres developed own scales for reporting the severity of side effects. Thus, it was nearly impossible to compare the outcome of clinical studies conducted at different centres in terms of normal tissue toxicity (Pavy et al., 1995). Therefore, a standardisation of the scoring procedure was (and still is) an important task, especially when conducting prospective multi-centre trials. To achieve uniform reporting of side effects different scoring systems defining a multitude of possible injuries have been developed and are regularly updated, especially by the Radiation Therapy Oncology Group (RTOG) and EORTC.

Proliferating tissues, such as epithelial tissues, are the first to respond to radiation dose leading to normal tissue reactions during or shortly after RT. Side effects that occur up to 90 days after treatment are referred to as *early (acute) side effects*. They usually heal completely, but may also turn into consequential late effects in case of very severe acute toxicities. Non-proliferating

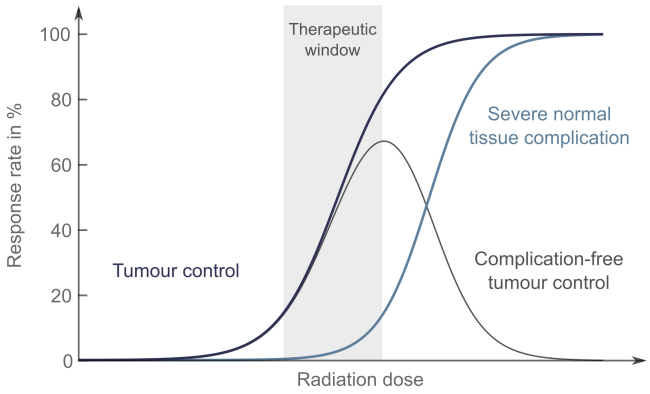


Figure 2.6: Schematic dose-response curves for tumour control and severe normal tissue complication. Adapted from Karger (2018).

tissues may react months or years after RT leading to *late (chronic) side effects*. These injuries are often permanent or even increase in their severity leading to serious organ dysfunction. Especially for patients with a long life expectancy, chronic side effects play an important role. Consequently, they are often chosen as study endpoints in clinical trials. In 1995, the late effects of normal tissues – subjective, objective, management and analytic (LENT-SOMA) scale has been developed to standardise the reporting of late side effects (Pavy et al., 1995; Rubin et al., 1995). Generally, the scoring systems are used by physicians to assess the dysfunction of more or less single organs.

In this work, the widely used scoring system Common Terminology Criteria for Adverse Events (CTCAE) version 4.0 by the US National Cancer Institute (2009) was employed. This scale covers a range of five grades of severity: mild (1), moderate (2), severe (3), life-threatening (4) and death (5). A revised version 5.0 has been published by the National Cancer Institute (2017).

Side effects following cranial irradiation Early side effects following cranial irradiation include rather unspecific reactions such as fatigue, somnolence, headache, nausea or vomiting and less frequent vertigo and seizures (Lawrence et al., 2010). Fatigue may be associated with irradiation of the brain stem (Ferris et al., 2018) and the cerebellum (Gulliford et al., 2012b) and simultaneous CTx, even though the mechanism of fatigue remains very often unknown. These symptoms are transient and can be treated with pharmaceutical drugs. Skin reactions such as alopecia (complete or partial hair loss) and erythema are also common (Salvo et al., 2010). Late side effects comprise radiation-induced brain necrosis (Burman et al., 1991; Shih et al., 2009; Lawrence et al., 2010), cranial neuropathies (Meeks et al., 2000; Shih et al., 2009) and neurocognitive dysfunction including memory and concentration impairment (Lawrence et al., 2010). Organ-specific toxicities such as otitis, tinnitus and hearing loss (Burman et al., 1991; Jereczek-Fossa et al., 2003; De Marzi et al., 2015; Lee et al., 2015), cataract (Kleiman, 2012), dry eye syndrome (Parsons et al., 1994), endocrine dysfunction (Constine et al., 1993; Pai et al., 2001; De Marzi et al., 2015) and visual impairment (Burman et al., 1991; Martel et al., 1997; Brizel et al., 1999), highly

depend on the tumour location. Table 2.1 summarises side effects and potentially associated OARs.

2.3.2 Patient-reported outcome measures – Quality of life

Physician-rated assessment of side effects aims to report essential organ dysfunctions objectively. The individual patient may, however, measure a successful treatment result based on further criteria, such as the ability to work, sleep and conduct daily activities, social participation or even outward appearance. Therefore, patient-reported outcome (PRO) measures that go beyond pure organ functions are becoming increasingly important in clinical studies, as they provide a more comprehensive picture of the success of the treatment in terms of patients' quality of life. There are also efforts to combine the physician-rated symptoms using the CTCAE with the patients' point of view in so-called PRO-CTCAE™ (Basch et al., 2014).

A multitude of quality of life questionnaires (QLQs) have been developed for various diseases and even different types of cancer or tumour location, for example, the Functional Assessment of Cancer Therapy (FACT) (Cella et al., 1993), Rotterdam Symptom Checklist (de Haes et al., 1990) or Symptom Distress Scale (McCorkle and Young, 1978). The Quality of Life Group of the EORTC has developed an integrated instrument for assessing QoL for cancer patients participating in clinical trials. The quality of life core questionnaire (QLQ-C30) addresses issues that patients with all types of cancer may experience. The supplementary modules focus on specific symptoms and functions for different tumour entities. The EORTC QLQs including symptom and functional scales have been thoroughly tested and validated and are available in different languages (Bottomley et al., 2005). The EORTC questionnaires employed in this thesis are presented in the following.

Core questionnaire QLQ-C30 This questionnaire comprises 30 questions that are combined into five functional scales, nine symptom scales and global health status. The functional scales address physical, role, social, emotional and neurocognitive function. The symptoms are fatigue,

Table 2.1: Side effects following cranial radiotherapy and potentially associated organs at risk.

Side effect	Associated organs at risk
Alopecia	Hair follicle within the skin
Cataract	Eye lenses
Cranial nerve failure	Associated cranial nerve
Dry eye syndrome	Lacrimal glands
Erythema	Skin
Fatigue	Brain stem, cerebellum
Hearing impairment	Cochlea, inner ear
Hypopituitarism	Pituitary
Memory impairment	Hippocampus, potentially further brain areas
Nausea	Brain stem or brain tissue
Radiation-induced brain necrosis	Brain stem of brain tissue

nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties (Aaronson et al., 1993), see appendix A table A.1.

Brain tumour module QLQ-BN20 The brain tumour-specific quality of life questionnaire (QLQ-BN20) only includes symptom scales that are specific for brain tumour patients: future uncertainty, visual disorder, motor dysfunction, communication deficit, headaches, seizures, drowsiness, itchy skin, hair loss, weakness of legs and bladder control (Taphoorn et al., 2010), see appendix A table A.2.

Analysis of quality of life scales A QLQ comprises various questions, so-called items. For evaluation, these questions are translated into symptom and functional scales. For some of these scales, only one single question is used (single-item measures), while for others several questions are combined to form a single score (multi-item measures). All questions are only used once on different scales. The scoring of the scales is described by Fayers et al. (2001). First, the raw score RS is calculated as the average of all n items I_i that together form one scale

$$RS = \frac{\sum_{i=1}^{n} I_i}{n}. \tag{2.6}$$

The raw score is then standardised to the final score S using a linear transformation to the interval [0, 100]. On symptom scales, high scores represent a high level of symptomatology. High scores for functional scales, however, represent a healthy level of functionality. This is taken into account when calculating the score S

$$\text{Functional scales} \qquad S = 100 \cdot \left(1 - \frac{RS - 1}{range}\right), \tag{2.7}$$

$$\text{Symptom scales} \qquad S = 100 \cdot \left(\frac{RS - 1}{range}\right), \tag{2.8}$$

in which *range* is defined as the difference between the minimum and maximum possible value of RS. For most items, *range* = 3 as they are scored from 1 to 4 with the following meaning: 1 *not at all*, 2 *a little*, 3 *quite a bit* and 4 *very much*. Global health status includes 7-point questions with *range* = 6.

Some patients may avoid answering questions. To handle such missing items, the *EORTC QLQ-C30 Scoring Manual* suggests the following: If at least half of the items have been answered a score is calculated from all available items, otherwise it is set to missing (Fayers et al., 2001).

Interpretation of quality of life scores To interpret the scores of the above-mentioned EORTC QLQ they can be compared to published data from healthy populations or other cancer patients. Examples are the *EORTC QLQ-C30 Reference Values Manual* (Scott et al., 2008) or data from a large Swedish population (Michelson et al., 2000). The clinical relevance of score differences

between patient groups or changes over time can be classified into *no change*, *a little*, *moderate* and *very much* for absolute differences in QoL scores of 0.0 – 4.9, 5.0 – 9.9, 10.0 – 19.9 and $\geq$ 20.0, respectively (Osoba et al., 1998).

2.3.3 Measures of neurocognitive function

Neurocognitive dysfunction is a frequent treatment or tumour sequela among long-term brain tumour survivors, such as patients with benign tumours or LGG with prolonged survival (Buckner et al., 2016) and young median age at diagnosis (Béhin and Delattre, 2003). Also for HGG patients, who have a short life expectancy, an important goal of treatment is the preservation of high QoL (Macdonald et al., 2005). Thus, several clinical trials in brain tumour patients include the assessment of neurocognitive function as a relevant outcome measure (Taphoorn and Klein, 2004; Correa, 2010; Schagen et al., 2014).

Neurocognitive sequelae may be caused by a variety of tumour-, patient- or treatment-related factors. These may also act synergistically although the exact mechanisms of neurocognitive dysfunction are not yet fully understood (Klein et al., 2002; Klein et al., 2003; Taphoorn and Klein, 2004). According to the interval between RT and the onset of CNS injury, neurocognitive dysfunction can be classified into acute, early delayed (subacute) and late delayed (Correa, 2010; Makale et al., 2017), see figure 2.7.

Several studies could not demonstrate an association between neurocognitive deficits and XRT (Armstrong et al., 2002; Klein et al., 2002). Other comparative studies on irradiated and non-irradiated LGG patients observed neurocognitive deficits (Surma-aho et al., 2001) as well as a decline in memory (Correa et al., 2007; Armstrong et al., 2012) or in attentional and executive

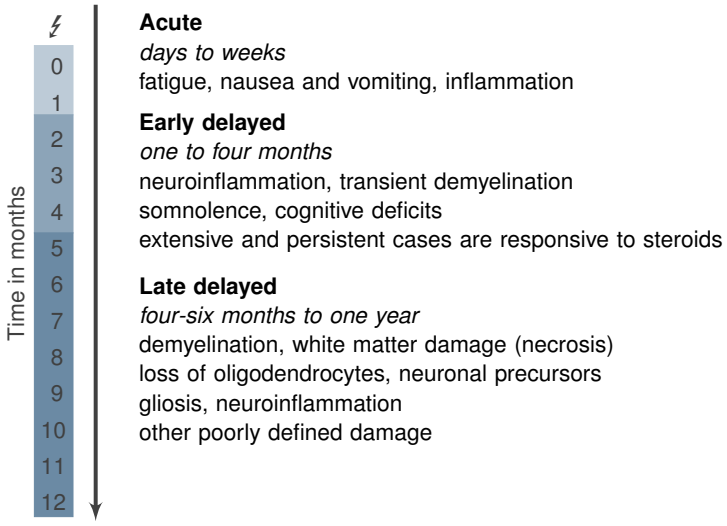


Figure 2.7: Time course of radiation-induced neurocognitive dysfunction. Adapted from Makale et al. (2017).

functioning (Douw et al., 2009). Both, radiation dose and volume of healthy brain tissue exposed to radiation, appear to be determining factors of neurocognitive function (Brown et al., 2003; Laack and Brown, 2004). In particular, the irradiation of certain structures of the brain such as the cerebellum (Merchant et al., 2014; Eekers et al., 2017) or the hippocampus (Gondi et al., 2012) may cause neurocognitive dysfunction. Thus, new RT techniques have been developed to spare these structures (Gondi et al., 2014; Kazda et al., 2014; Tsai et al., 2015; Zhao et al., 2017).

So far, most studies on neurocognitive function following PBT have been conducted in paediatric populations that generally showed neurocognitive preservation (Greenberger et al., 2014; Kahalley et al., 2016; Yang et al., 2019) or low decline in certain patient subgroups, for example, young children (Greenberger et al., 2014), patients receiving high doses to the left hippocampus (Greenberger et al., 2014; Zureick et al., 2018) and patients treated with craniospinal irradiation (Pulsifer et al., 2018). Studies describing the neurocognitive outcome of adult brain tumour patients treated with protons are rare. In a small cohort of 20 adult LGG patients treated with PBT no evidence for an overall decline in neurocognitive function or QoL could be observed (Shih et al., 2015; Sherman et al., 2016).

Although neurocognitive dysfunction is an important side effect, it is difficult to measure. Along with extensive neuropsychological assessment, a variety of rather objective and standardised test batteries for neurocognitive outcome have been developed (van Loon et al., 2015). A commonly used test is the Mini-Mental State Examination (MMSE), as it requires only a short time to complete (Folstein et al., 1975). Among brief screening instruments, the Montreal Cognitive Assessment (MoCA) test (Nasreddine et al., 2005) showed superior sensitivity in detecting neurocognitive impairment compared to the MMSE, 61.9 % vs 19.0 % at pre-defined cut-off scores (Olson et al., 2011a). Since the MoCA test was analysed in this thesis, it is briefly described below.

The Montreal Cognitive Assessment test The MoCA test was developed as a brief screening tool to detect mild cognitive impairment (Nasreddine et al., 2005). Nevertheless, the MoCA test has already been used in patients with brain tumours (Olson et al., 2008; Olson et al., 2011b). It is available in both paper-and-pencil and electronic format and in several languages. Thus, it is applicable regardless of nationality and allows for comparable results over diverse studies. The MoCA test includes different tasks assessing several neurocognitive domains (visuospatial, executive, naming, attention, concentration, calculation, language, abstraction, short-term memory, orientation in time and space). A more detailed explanation of the individual tasks for each domain is given in appendix A. The single scores of each test sum up to a total score of 30 points. In trials with repeated measurement, learning effects may occur that may compromise the assessment of neurocognitive function. To minimise these learning effects, three different test versions have been developed (Cooley et al., 2015).

The authors of the MoCA test initially suggested that below a test score of 26 mild cognitive impairment is possible. This cut-off value has been discussed and may vary by age, education level and ethnicity (Gagnon et al., 2013; Milani et al., 2018; Thomann et al., 2018).

2.4 Normal tissue complication probability models

An essential part of treatment plan assessment is to check whether the present dose distribution exceeds tolerance doses of affected OARs to avoid severe side effects, see section 2.2.5. These tolerance doses are results of clinical experience and information about the dose to different organs, first summarised by Emami et al. (1991). This seminal work has been further developed by numerous studies investigating the relationship between dose distribution and side effects. These dose-response relationships can be expressed using various NTCP models. Some of the different types of NTCP models are presented below.

2.4.1 Types of NTCP models

NTCP models aim to predict the probability of complications based on the dose distribution in associated irradiated organs. For this, the real three-dimensional dose distribution is reduced to a few simple metrics. Different methods for modelling clinical outcome data of retrospective patient cohorts and their dose distributions are described in the following (Gulliford, 2015).

DVH-reduction models Based on the data published by Emami et al. (1991), the empirical Lyman-Kutcher-Burman model (LKB model) was developed that describes the dose-response as a function of irradiated volume, reduces DVH to a single metric and estimates model parameters for specific OARs (Lyman, 1985; Kutcher and Burman, 1989; Burman et al., 1991; Kutcher et al., 1991). The model parameters are TD_{50}, m and n. The parameter $TD_{50}(V)$ is the tolerance dose for uniform irradiation of a partial volume V of an OAR at which 50 % of patients are likely to experience toxicity. The parameter m represents the slope at the steepest part of the dose-response curve, see figure 2.6. n indicates the volume effect of the investigated OAR (Gulliford et al., 2012b). Serially structured organs such as the spinal cord are described by $n \approx 0$, while parallel organs are characterised by $n \approx 1$. Taking fractional irradiation into account, the normal

tissue complication probability for an LKB model $NTCP_{\mathrm{LKB}}$ for a uniform dose D to a volume V of an OAR is given by

$$NTCP_{\mathrm{LKB}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(\frac{-u^2}{2}\right) \mathrm{d}u, \tag{2.9}$$

$$\text{where } t = \frac{D - TD_{50}(V)}{m \cdot TD_{50}(V)}, \tag{2.10}$$

$$TD_{50}(V) = \frac{TD_{50}(V_{\mathrm{OAR}})}{V^n}, \tag{2.11}$$

where V_{OAR} represents the entire volume of the considered OAR.

However, dose distributions to OARs are actually non-uniform. The inhomogeneous dose distribution can be reduced to a single metric that produces the same probability of a given side effect as a corresponding uniform dose distribution. Such a metric is the widely used generalised equivalent uniform dose (*gEUD*) given by

$$gEUD = \left(\sum_i V_i D_i^a\right)^{1/a}, \tag{2.12}$$

where D_i is the dose defined for each bin i in a differential DVH, see figure 2.5B. V_i is the volume in a dose bin i and a is a volume parameter that is equivalent to $1/n$. This calculated uniform dose can then be applied as $D = gEUD$ in the LKB model in equation (2.10).

Tissue-architecture models Models have been developed that consider the functional architecture of the tissue by introducing functional subunits of an OAR. These can be anatomical substructures, such as nephrons of the kidney, or the largest cell group that still functions as long as it comprises a surviving clonogen (Gulliford, 2015). These functional subunits can be arranged in serial or parallel order, or a combination of both. In parallel organs, functional subunits are performing rather independently so that side effects occur after the irradiated volume exceeds a critical value. Side effects that arise from irradiation of parallel organs rather depend on the mean dose deposited in these organs (e.g. liver, lung or kidney). This effect is described in an LKB model with a parameter $n \approx 1$. Serially structured organs, such as the spinal cord, lose their entire organ function when a small area is irradiated with high doses. An LKB model characterises such an OAR with the parameter $n \approx 0$.

Källman et al. (1992) suggested the *relative seriality model* and Niemierko and Goitein (1993) presented a *critical volume model* based on the assumption that NTCP is fully determined by the number of surviving functional subunits of an OAR.

Multiple-metric models The above-mentioned models predict the complication probability for one specific side effect based on the dose to a corresponding OAR. However, some complications are affected by the irradiation of different OARs, e.g. swallowing dysfunction following the irradiation of superior pharyngeal constrictor muscle and the supraglottic larynx (Christianen et al., 2012) or heart valvular dysfunction by the irradiation of heart and lung (Cella et al., 2014). To correct for this in LKB models, an interaction *gEUD* variable for both OARs can be introduced (Cella et al., 2014). Nevertheless, side effects may also be modified by dose-independent clinical parameters, such as age or radiation technique (Christianen et al., 2012). Multivariable logistic regression models are appropriate to include clinical and dosimetric parameters. They are defined by

$$NTCP_{\text{Logistic}} = \left(1 + \mathrm{e}^{-g(x)}\right)^{-1}, \tag{2.13}$$

$$\text{where } g(x) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i, \tag{2.14}$$

where β_i denote model coefficients and x_i are p individual explanatory variables.

A probability of 50 % for the occurrence of the considered side effect occurs if the condition

$$g(x) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i, = 0 \tag{2.15}$$

is met. For univariable models including a single dosimetric parameter as a predictor, TD_{50} is calculated by

$$TD_{50} = -\frac{\beta_0}{\beta_1}. \tag{2.16}$$

Models considering the development over time For the analysis of longitudinal toxicity data with repeated measures, as usual in patient follow-up, a generalised estimating equation (GEE) approach can be applied (Liang and Zeger, 1986). For a cohort consisting of N patients, each patient is seen for T follow-up visits. This number of visits T may differ between patients (Agresti, 2002). An observation for patient $i \in [1, N]$ at follow-up visit t can be expressed by y_{it}. Thus, all observations of this patient over time form the vector $Y_i = (y_{i1}, ..., y_{it}, ..., y_{iT})^{\mathsf{T}}$, $t \in [1, T]$. The K corresponding explanatory variables for patient i at time t are combined in the vector $\boldsymbol{x}_{it}$. The value of variable k is noted with $x_{itk}, k \in [0, K]$. For $k = 0$, $x_{it} = 1$ for all patients. As the side effects are measured repeatedly for each patient at different follow-up visits, these observations are assumed to be correlated, but independent across different patients (Agresti, 2002; Wang, 2014). An appropriate link function $g(.)$ relates the explanatory variables to the expected value $E(y_{it}) = \mu_{it}$ via

$$g(\mu_{it}) = \boldsymbol{x}_{it}^{\mathsf{T}} \beta, \tag{2.17}$$

where β is an unknown vector containing all K regression coefficients (Samur et al., 2014). For the evaluation of side effects, the severity grades are commonly dichotomised, e.g. grade < 2 vs grade ≥ 2. For such binary outcome data, y_{it} is either 0 or 1. In this case, the logit function may be an appropriate link function (Samur et al., 2014).

The regression coefficients β_k can be estimated by the solution (e.g. using a maximum likelihood approach) of the GEE

$$\sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta_k} V_i^{-1} (Y_i - \mu_i) = 0, k \in [0, K], \tag{2.18}$$

where V_i indicates the variance-covariance matrix for Y_i of patient i. It is expressed by

$$V_i = \phi A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}, \tag{2.19}$$

where A_i is a $T \times T$ diagonal matrix with the variance of Y_i as the t-th diagonal element (Liang and Zeger, 1986), ϕ is a scale parameter depending on the distributions of outcomes and $R_i(\alpha)$ indexed by a vector of association parameters α represents the *working* correlation matrix of size $T \times T$. The *working* correlation matrix models the dependence of each observation with other observations for the same patient. The elements of this matrix are the correlations between the longitudinal observations within a patient. Commonly used working correlation structures for GEE models are for example independent, autoregressive or unstructured (Wang, 2014). The advantage of GEE models is that consistent estimates of the parameters can be obtained even if the *working* correlation matrix is incorrectly specified (Samur et al., 2014).

2.4.2 Endpoint definition and parameter fitting

Data on side effects should be prospectively collected in clinical trials according to standardised tests or grading systems for a defined follow-up duration, see section 2.3. For modelling, these grades or scores may then be dichotomised according to severity (e.g. grade < 2 vs grade ≥ 2). During univariable modelling, the eligible parameters are preselected. A method for fitting the model parameters is maximum likelihood estimation (Cella et al., 2014; Gulliford, 2015). To find the optimal coefficients, the agreement of the predicted outcome with the observed outcome is maximised. For multivariable logistic regression models, the individual variables should not be strongly correlated. This should be considered when selecting the appropriate model predictors. DVH parameters are often correlated within a given cohort that may lead to problems with multicollinearity (Bentzen et al., 2010). Principle component analysis may be a method to overcome this problem (Dawson et al., 2005). However, the principal components consist of various dosimetric parameters, which make them appear abstract and are therefore less practical in clinical routine (e.g. in treatment planning). Cross-validation and bootstrapping ensure generalisability (Gulliford, 2015).

2.4.3 Assessment of model performance

Model performance can be assessed in terms of discrimination and calibration. Discrimination refers to the ability to separate patients who do or do not develop a given side effect. Calibration compares the predicted and observed outcome (Bentzen et al., 2010).

The discriminating ability of NTCP models can be assessed using receiver-operating characteristic (ROC) analysis by calculating sensitivity and specificity (Gulliford, 2015). Sensitivity and specificity are calculated from the contingency table of the predicted and observed outcome for each possible cut-off of the continuous model variable. The ROC curve is a plot of sensitivity against 1-specificity. The area under the receiver operating characteristic curve (*AUC*) is a metric that ranges from 0 to 1. A value of 0.5 represents a random prediction, a value of 1 a perfect prediction. In the medical field, *AUC* values do rarely exceed a value of 0.8 (Gulliford, 2015).

In a calibration plot, the observed outcome is plotted against the predicted NTCP values. For perfectly calibrated models, the data points are aligned on the quadrant bisector. Calibration is important if the models shall predict exact complication probabilities, e.g. for comparison between different treatment plans.

2.4.4 Model validation

In this thesis, NTCP models are developed on a training cohort and internally validated using cross-validation. Model performance has to be checked by applying the final NTCP model with unchanged model coefficients on an external patient cohort and calculating *AUC* and the calibration plot. The Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Initiative published recommendations for the reporting of development and validation of prediction models (Moons et al., 2015). The classification of different analysis types is presented in appendix A table A.3. Testing on an independent dataset from another institution (external validation) may reduce model performance due to different scoring of side effects, patient demographics and comorbidities or treatment strategies (Bentzen et al., 2010).

2.5 Model-based approach for patient selection for proton beam therapy

The number of operating proton facilities is increasing. The high technical and time expenditure leads to high costs of this treatment modality. These high costs and their uncertain reimbursement by health insurance companies limit the capacity for PBT. Therefore, it is important to offer this treatment to those patients who may benefit most from PBT compared to XRT. However, there are only a few RCTs that directly compare XRT and PBT, as such studies are complex to conduct for several reasons (Liao et al., 2018). A feasible approach to meet these challenges and to identify patients suitable for PBT is the so-called *model-based approach* (Langendijk et al.,

2013). In the Netherlands and Denmark, it has already been implemented in clinical practice for patients with head and neck cancer. This section discusses the need for alternatives to RCTs and the principles of the two-phase model-based approach as proposed by Langendijk et al. (2013).

2.5.1 Limits of randomised controlled trials

Decisions in healthcare policy and practice are mainly evidence-based. Systematic reviews of RCTs are considered the highest level of scientific evidence to explain a cause-effect relationship between treatment and outcome (Burns et al., 2011). In an RCT, patients are randomly assigned to either a treatment or control group. Ideally, neither patients nor clinicians should know the character of intervention (double-blind study) (Sibbald and Roland, 1998). These properties make RCTs the gold standard, e.g. in pharmaceutical research, but they also have disadvantages, especially in practical and ethical aspects. RCTs are usually expensive and time-consuming. Blinding conditions are not always feasible and reimbursement of costs of experimental treatments cannot be guaranteed. Ethical concerns may arise, for example, if patients shall be treated only with a placebo or a presumably less effective therapy.

In contrast to RCTs when comparing systemic therapies, there are challenges in performing RCTs in RT to compare different techniques and draw general conclusions. When comparing a new with a conventional irradiation technique, there are many degrees of freedom which may vary between different institutions. There exists no *standard application* of RT, neither XRT (Fairchild et al., 2014; Das et al., 2017) nor PBT (Taylor et al., 2016). Both terms are very general as they cover a wide variety of techniques (Widder et al., 2016). The heterogeneity between centres in terms of treatment planning systems, quality assurance, training skills, image guidance, treatment adaptation, immobilisation strategies etc. may be so pronounced that it may be difficult to transfer results from RCTs into clinical routine (Widder et al., 2016; Langendijk et al., 2018). Hence, alternative approaches need to be applied to generate evidence.

The dose distribution characteristics of PBT compared to XRT allow for two different treatment strategies. On the one hand, dose-escalation in the target region can be achieved without additional exposure to adjacent normal tissues. Dose escalations to increase tumour control should be investigated in the context of RCTs, as it is not known whether tumour control actually improves and to what extent side effects are induced (Langendijk et al., 2018). On the other hand, healthy tissue can be spared to a higher extent than in XRT while achieving equivalent target coverage. Comparing PBT and XRT in terms of reduced side effects, some obstacles make RCTs difficult to perform. Late radiation-induced side effects as relevant endpoints of such trials may manifest many years following RT. Thus, results from large long-term RCTs may be obsolete as RT (and PBT in particular) is still a rapidly evolving technology (Langendijk et al., 2013). Usually, all eligible patients are enrolled in an RCT, regardless of their presumed extent of side effect reduction under PBT. However, the expected benefits from PBT may vary largely between patients. An RCT may answer the question whether PBT is generally superior to XRT in terms of reduced

side effects for certain tumour entities and patient groups, but cannot provide an answer about the clinical benefits at a patient-individual level. Thus, RCT designs could address the heterogeneous extents of treatment effects by the inclusion of NTCP models (Scherman et al., 2019). Furthermore, to observe a significant difference between both treatment modalities in a small number of patients and thus a short study duration, a proper preselection of eligible patients is necessary. Due to this preselection process, the estimated difference in complication probability between both treatment groups is enhanced, leading to a supposed lack of equipoise. As a result, RCTs can be considered unethical (Bentzen, 2008; Goitein and Cox, 2008; Suit et al., 2008).

2.5.2 Principles of the model-based approach

The model-based approach consists of two phases: model-based selection and model-based validation. The first phase, in turn, comprises three steps: development and validation of NTCP models, patient-specific plan comparison and estimation of the clinical benefit of PBT. The individual steps are explained in more detail below.

Phase α: Model-based selection

Patients shall be selected regarding their reduction of side-effect probabilities under PBT compared to XRT. If this reduction exceeds a given threshold, those patients will be suitable for PBT treatment. The side-effect probabilities for each patient are estimated using NTCP models.

1. Development and validation of NTCP models The relation between dose to OARs and associated side effects is mathematically described by NTCP models. For different entities and relevant side effects that may occur following RT, multivariable NTCP models have to be developed and externally validated. The values of dosimetric parameters that are supposed to be important in these models should be reduced, if possible, during treatment planning. Development and external validation of NTCP models are described in section 2.4.

Up to now, most NTCP models have been derived from data of patients treated with XRT. It has been demonstrated that NTCP models can differ between various XRT techniques (Beetz et al., 2012b; Christianen et al., 2012; Troeller et al., 2015). Since dose distributions of XRT and PBT differ substantially, the same side effect may be related to different dosimetric parameters. If photon and proton models differ too much for certain side effects, technique-specific NTCP models may become necessary.

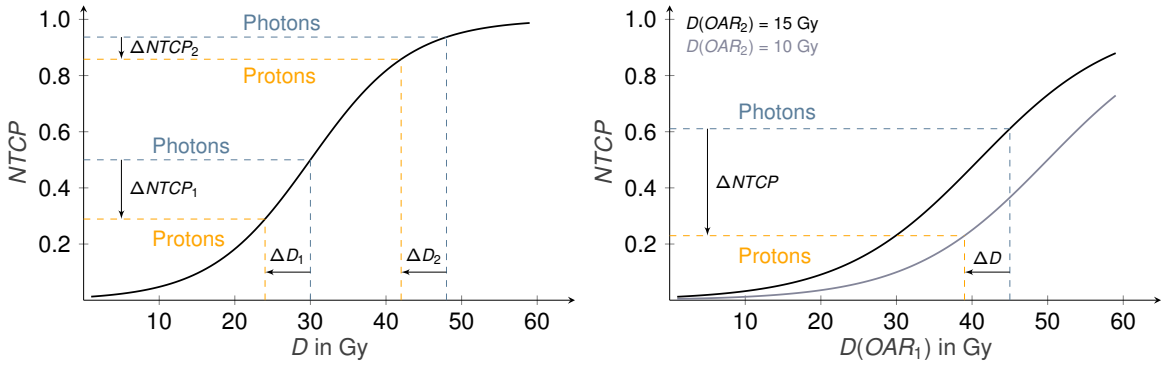
To use existing photon-derived NTCP models to predict the NTCP for PBT, these models need to be validated on prospectively collected PBT patient data. Blanchard et al. (2016) succeeded to validate established NTCP models on head and neck cancer patients treated with PBT. However, Pedersen et al. (2020) observed large differences in the prediction of rectal complications in prostate cancer patients between both treatment modalities. Continuous model validation and

updating may be implemented, for example, in the framework of a rapid learning health care system (Lambin et al., 2013a; Lambin et al., 2015).

2. Individual *in silico* planning comparative studies For each patient, two treatment plans are created, one with protons and the other with a state-of-the-art photon RT technique. The dose distributions of both treatment plans are compared, especially the dosimetric parameters of OARs that are predictors in the NTCP models.

3. Estimation of the clinical benefit The treatment plans and NTCP models are used to estimate the difference between side-effect probabilities of PBT and XRT. Figure 2.8A shows impressively that an equal dose difference between a photon and a proton treatment plan may lead to different NTCP reductions. This effect depends on the slope of the NTCP curve in which both dose values are located. In a multivariable model including dosimetric predictors of two different OARs, the difference in NTCP between the proton and photon plan may be even higher if PBT is able to spare both OARs simultaneously (Christianen et al., 2012), see figure 2.8B. In some cases, the occurrence of certain side effects is mainly determined by clinical cofactors (e.g. comorbidities, administration of CTx). Hence, the dose difference may play a minor part in the occurrence of the side effect (Lambin et al., 2013b).

A patient is finally selected for PBT if the extent of NTCP reduction in the PBT plan compared to XRT exceeds a given threshold. This predefined threshold depends on the severity of the considered side effects, with lower thresholds for more severe toxicities. For toxicities of CTCAE grade 2, 3 and 4 – 5, the Dutch Society of Radiation Oncology suggests thresholds of 10, 5 and 2 percentage points, respectively (Langendijk et al., 2018). In some cases, multiple side effects are considered in the selection procedure (NTCP profiles). Here, both the NTCP difference of



(A) Univariable NTCP model.

(B) Multivariable NTCP model.

Figure 2.8: Model-based approach according to Langendijk et al. (2013). **(A)** Two equal dose differences in an OAR between a photon and proton plan ($\Delta D_1 = \Delta D_2$) translate into different reductions in NTCP ($\Delta NTCP_1 > \Delta NTCP_2$). **(B)** In a multivariable model including dosimetric predictors of two different OARs, the NTCP difference between proton and photon plan may be even higher if PBT is able to reduce dose to both OARs simultaneously.

every single endpoint as well as the summarised NTCP difference for all considered endpoints must exceed different thresholds (Langendijk et al., 2018). If the NTCP difference remains below the recommended threshold, the patient is treated with state-of-the-art XRT.

Phase β: Model-based validation

The initial hypothesis of reduced side effects after PBT compared to XRT is evaluated during model-based validation, see figure 2.9. Patients who were selected for PBT during phase α are enrolled in such prospective validation studies and treated with the proton treatment plan created during step 2. The calculated NTCP values for both treatment modalities are averaged over all patients enrolled in the study to estimate the expected toxicity rates. The finally observed toxicity rates of patients treated with PBT are then compared to the initially predicted proton NTCP values to detect possible shortcomings of the applied NTCP model (Langendijk et al., 2018). Furthermore, it can be tested whether the observed toxicity rate following PBT is indeed lower than the estimated NTCP values for XRT (calibration-in-the-large (Steyerberg et al., 2010)). However, the true toxicity rates after PBT are only compared with a supposed toxicity probability for XRT.

Another way to compare the real outcome of PBT and XRT directly is the use of prospectively collected patient data as control groups, e.g. historical cohorts or patients treated in radiotherapy

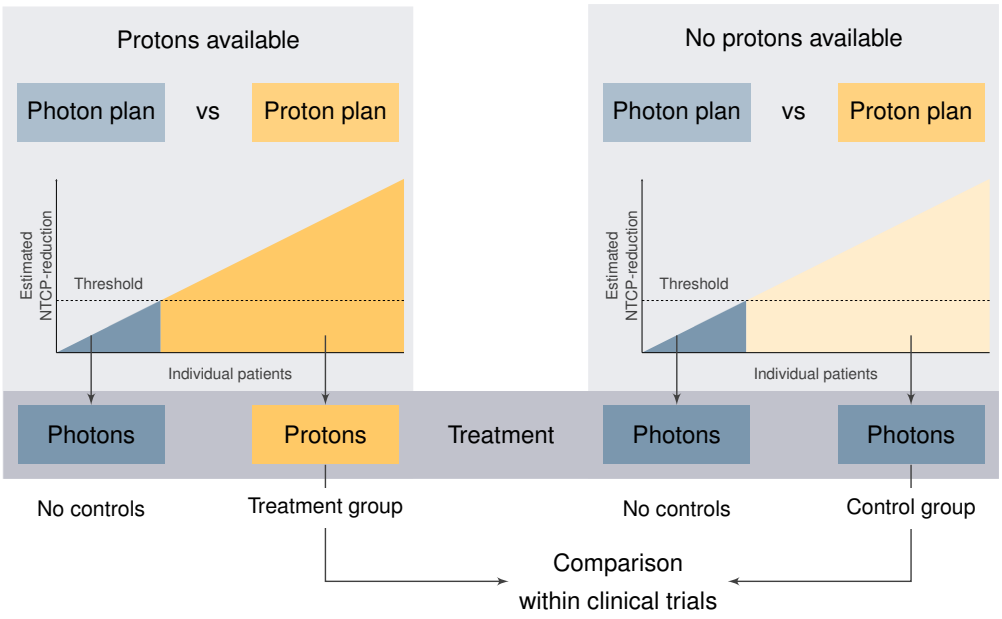


Figure 2.9: Schematic overview of the model-based validation according to Langendijk et al. (2013). Where PBT is available, patient selection according to phase α of the model-based approach is applied (left side). The outcome of these patients treated with PBT can be compared to prospectively collected data from patients routinely treated with photons (right side). To ensure a fair comparison between both groups, only those patients whose estimated NTCP reduction would have exceeded the threshold for PBT are included in the control group.

centres without access to PBT (Langendijk et al., 2013). Patients selected for and treated with PBT according to phase α are enrolled in such clinical trials as the treatment group, see figure 2.9. It is not reasonable to compare the outcome of this treatment group with the outcome of patients who were rejected in phase α and treated with XRT instead. Therefore, an independent control group has to be designed. It should consist of patients who would have been candidates for PBT according to the same selection procedure but were still treated with XRT (the control group in figure 2.9).

3 Investigated patient cohorts

In this thesis, the outcome of patients with intracranial tumours treated with PBT is investigated. Similar to what is proposed by the model-based approach (Langendijk et al., 2013), NTCP models for early and late side effects are developed and validated in chapter 4. Chapter 5 investigates dose-effect relations for neurocognitive function and QoL. Finally, a treatment plan comparison study is conducted in chapter 6. Based on these treatment plans and NTCP models, the potential benefit of PBT is estimated. In this chapter, all patients analysed in any of the following studies are presented and characterised.

Patient cohorts Patient cohorts from three different PBT centres were available. These three institutes are the University Proton Therapy Dresden (UPTD) at the Department of Radiotherapy and Radiation Oncology of the University Hospital Carl Gustav Carus Dresden, the West German Proton Therapy Centre Essen (WPE) and the Massachusetts General Hospital (MGH) in Boston. Patients treated at UPTD were enrolled in different clinical trials (DRKS00007670, 2015; DRKS00008569, 2015; NCT02824731, 2016) and treated using a DS technique. Patients from WPE were enrolled in the clinical trial (DRKS00004384, 2012) and treated using a PBS technique. At MGH, patients were irradiated with a DS technique. All patients were consecutively treated between 2013 and 2017.

General inclusion criteria were age $\geq$ 18 years, diagnosis of a tumour in the brain or skull base, normo-fractionated PBT, Eastern Co-operative of Oncology Group (ECOG) performance status $\leq$ 2 as well as written informed consent. Exclusion criteria were inability to MRI scans, lack of patient compliance and limited possibilities for reproducible positioning. All studies were approved by the local Ethics Committee and the respective institutional review boards.

Not all patients could be included in every study described in this thesis, see figure 3.1. For example, neurocognitive testing was available only for patients who were enrolled in the clinical trial Proto-R-Hirn (DRKS00007670, 2015). Moreover, for patients, that were enrolled in the ongoing clinical trial Proto-Choice-Hirn (NCT02824731, 2016), late toxicity data could not be evaluated as this is the primary endpoint of this study. Consequently, the patient composition will change for the analyses presented in the following chapters.

Patient characteristics Table 3.1 gives an overview of several patient characteristics. Generally, the patient cohorts differed in most characteristics, especially in tumour histologies. While the majority of patients treated at UPTD had glioblastomas (29 %), only 11 % of patients treated at WPE were diagnosed with glioblastoma. In the cohort treated at MGH, there was no patient with glioblastoma at all. In both WPE and MGH cohort, patients with meningioma were most frequently present (37 % and 40 %, respectively). The different composition of the cohorts in terms

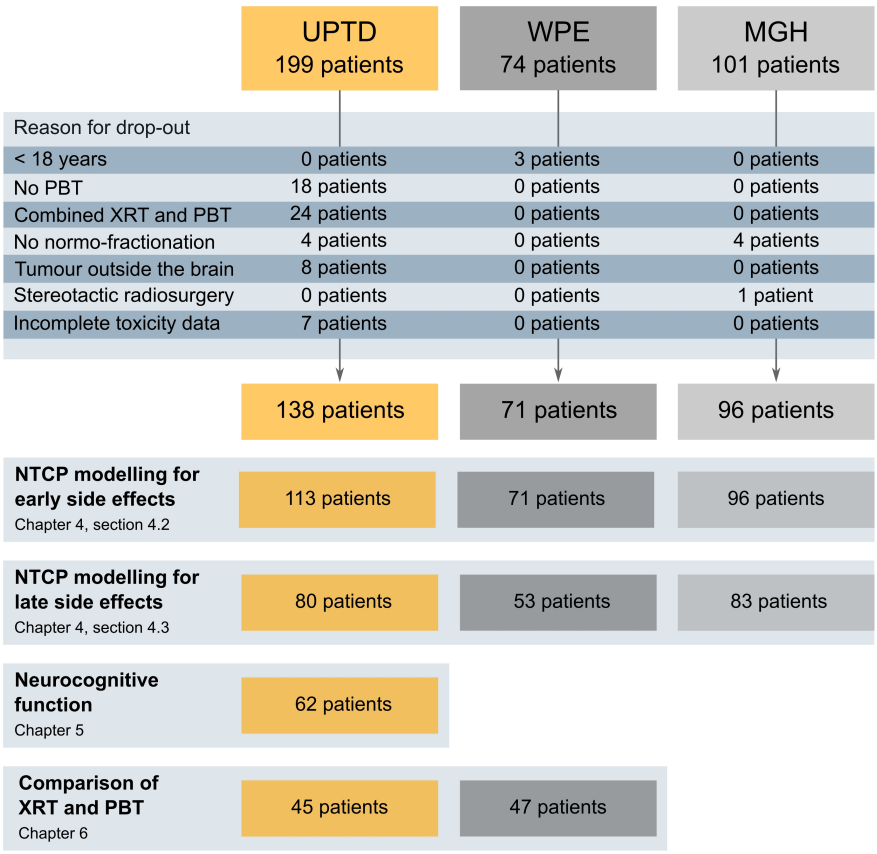


Figure 3.1: Investigated patient cohorts.

of tumour histologies results in differences concerning surgery and CTx. In the UPTD cohort, more patients underwent surgery or additional CTx compared to the WPE and MGH cohorts. Tumour volumes were considerably larger for patients irradiated at UPTD than for patients of the other two centres. Furthermore, the prescribed dose with a median value of 60 Gy(RBE) was also higher in this cohort compared to the WPE and MGH cohort, which had a median prescribed dose of 54 Gy(RBE). The median age was slightly higher in the UPTD cohort (49.9 years) than in the other two groups (45.1 years and 46.1 years). Slightly more men than women were treated in the UPTD cohort, while the ratio was almost reversed in the other two groups. Most patients in the WPE and UPTD cohort had the tumour in the right hemisphere (39 % and 49 %, respectively), while in most MGH patients the tumour was located on the left side (39 %). In the WPE and UPTD cohorts, tumours affected the temporal lobe in most patients (59 % and 30 %, respectively), while this was observed in only 25 % of patients in the MGH cohort. No significant difference between the cohorts with regard to tumour location in the brain or skull base could be observed.

Table 3.1: Characteristics of the investigated patient cohorts. To compare continuous variables, a Kruskal-Wallis test was applied, for categorical variables, a χ^2 test was used.

	UPTD		**WPE**		**MGH**		
Number of patients	138		71		96		
PBT technique	DS		PBS		DS		
Characteristic	**Median**	**(Range)**	**Median**	**(Range)**	**Median**	**(Range)**	**p-value**
Age at PBT in years	49.9	(18.8 – 84.7)	45.1	(18.1 – 79.6)	46.1	(18.5 – 89.3)	0.022
Tumour volume (CTV) in cm^3	136.7	(1.2 – 499.1)	54.6	(2.1 – 368.1)	28.1	(1.2 – 267.2)	< 0.001
Prescribed dose in Gy(RBE)	60.0	(30.0 – 74.0)	54.0	(42.5 – 60.0)	54.0	(36.0 – 70.2)	< 0.001
	N	**%**	**N**	**%**	**N**	**%**	
Gender							0.018
Male/female	80/58	(58/42)	33/38	(46/54)	38/58	(40/60)	
Surgery							0.026
No/yes/missing	18/119/1	(13/86/1)	20/51/0	(28/72/0)	21/75/0	(22/78/0)	
Chemotherapy							0.011
No/yes/missing	84/53/1	(61/38/1)	55/16/0	(77/23/0)	74/22/0	(77/23/0)	
Re-irradiation							0.032
No/yes/missing	117/20/1	(85/14/1)	66/3/2	(93/4/3)	84/6/6	(88/6/6)	
Tumour location							0.19
Brain/skull base/other	76/61/1	(55/44/1)	37/32/2	(52/45/3)	39/55/2	(41/57/2)	
Tumour location							0.009
Left hemisphere	49	(36)	20	(28)	37	(39)	
Right hemisphere	68	(49)	28	(39)	27	(28)	
Central	18	(13)	20	(28)	28	(29)	
Bilateral	3	(2)	3	(4)	4	(4)	
Tumour location							< 0.001
Temporal lobe	42	(30)	42	(59)	24	(25)	
Frontal lobe	34	(25)	34	(48)	22	(23)	
Parietal lobe	6	(4)	6	(8)	7	(7)	
Occipital lobe	4	(3)	4	(6)	3	(3)	
Multiple lobes	30	(22)	30	(42)	4	(4)	
Other	22	(16)	22	(31)	36	(38)	
Histology*							< 0.001
Glioblastoma	40	(29)	8	(11)	0	(0)	
Astrocytoma III	8	(6)	12	(17)	10	(10)	
Oligoastrocytoma III	2	(1)	2	(3)	5	(5)	
Oligodendroglioma III	1	(1)	1	(1)	0	(0)	
Astrocytoma I-II	13	(9)	5	(7)	8	(8)	
Oligoastrocytoma I-II	6	(4)	2	(3)	2	(2)	
Oligodendroglioma I-II	5	(4)	1	(1)	1	(1)	
Ependymoma	1	(1)	1	(1)	2	(2)	
Craniopharyngioma	1	(1)	3	(4)	2	(2)	
Meningioma	16	(12)	26	(37)	38	(40)	
Medulloblastoma	0	(0)	2	(3)	0	(0)	
Pituitary adenoma	10	(7)	1	(1)	11	(11)	
Vestibullar schwannoma	0	(0)	0	(0)	9	(9)	
Other	35	(25)	7	(10)	8	(8)	

Abbreviations: N, number of patients; DS, double scattering; PBS, pencil beam scanning; CTV, clinical target volume; UPTD, University Proton Therapy Dresden; WPE, West German Proton Therapy Centre Essen; MGH, Massachusetts General Hospital. *Tumour classification according to WHO Classification of Tumours of the Central Nervous System, Fourth Edition.

4 Modelling of side effects following cranial proton beam therapy

Cranial RT may be associated with mild to severe side effects that can adversely affect patients' QoL. Patients with HGGs have a short life expectancy (Stupp et al., 2009) and thus, an important goal of treatment is the preservation of QoL as long as possible due to reduction of early side effects. Patients with benign tumours or LGGs often have prolonged survival and young age at diagnosis (Buckner et al., 2016). These long-term survivors may be more affected by late side effects.

Based on dosimetric and clinical outcome data, NTCP models for different endpoints have been developed for XRT (Burman et al., 1991; Bender, 2012; Gondi et al., 2012). However, studies based on data from PBT or combined PBT and XRT are rare (De Marzi et al., 2015; Palma et al., 2020). In this chapter, common physician-rated early and late side effects following cranial PBT were investigated in patient cohorts from three PBT centres. These data were used for developing NTCP models and for their external validation. Furthermore, an additional study was carried out that investigated the interobserver variability for the side effects erythema and alopecia.

Parts of the work presented in this chapter have been published (Dutz et al., 2019) and were presented at several conferences (Dutz et al., 2017a; Dutz et al., 2017b; Dutz et al., 2018).

4.1 Experimental design for modelling early and late side effects

The patient cohorts have been characterised comprehensively in chapter 3. The cohorts investigated in each analysis for early and late side effects are described separately in the respective section. They differed slightly since long-term follow-up was not available for all patients.

Radiotherapy planning and treatment delivery

For the patient cohorts treated at the UPTD and the MGH (except for two cases that were treated with PBS due to scheduling reasons), treatment plans were calculated using a DS technique and the treatment planning system XiO® (Elekta AB, Stockholm, Sweden). For the patient cohort treated at the WPE, a PBS technique using XiO® was applied for patients treated between 2013 and 2014 and RayStation® (RaySearch AB, Stockholm, Sweden) was used for patients treated between 2015 and 2016. For delineation and treatment planning, a CT scan with a slice thickness of 3 mm (UPTD), 1 mm (WPE) or 1.25 mm (MGH) was acquired for each patient and rigidly registered with (post-operative) MRI scans. The GTV included the tumour or the resection cavity

with residual tumour volume, if present. The CTV included microscopic disease and oedema in T2-weighted MRI scans taking into account anatomical boundaries.

In the cohort treated at UPTD, the target volume for the DS treatment plans was the CTV with an in-beam margin of 3.5 % of the distal or proximal CTV depth and an additional absolute margin of 2 mm. In the cohort treated at MGH, the distal range expansion to the CTV was 3.5 % and 1 mm with an additional modulation expansion of twice the distal expansion. For the cohort treated at WPE using a PBS technique, a PTV was constructed by adding a 5 mm margin to the CTV. For treatment planning, one to four beams in a patient individual beam setting were applied according to clinical practice in the respective centres. The RBE was considered to be constant at 1.1. Thus, the unit Gy(RBE) is used in the following. The dose-volume constraints differed slightly between the cohorts treated at UPTD and WPE, see appendix B tables B.4 and B.5. Patient individual planning objectives were used at MGH. During PBT, patient positioning was daily verified by manual registration of orthogonally paired X-ray images with digitally reconstructed radiographs of the planning CT. In the cohort treated at MGH, most patients were immobilised using a stereotactic bite block system.

Definition of endpoints and extraction of dose-volume parameters

Treatment-related side effects were prospectively scored at baseline, weekly during PBT, at the end of PBT, 3 months after PBT and thereafter in regular intervals depending on the institution. The CTCAE version 4 grading system was used. For early side effects, the maximum severity grade was assessed during treatment and up to 3 months after PBT. The long-term follow-up schemes differed between the institutions: every 3 months at UPTD, every 6 months at MGH and every 12 months at WPE. Thus, late side effects were evaluated at 12 and 24 months after PBT. For patients with non-zero baseline value and an increase in severity during treatment or follow-up, the maximum score was used regardless of the pre-treatment value otherwise grade 0 was used. All endpoints were dichotomised at grade $\geq$ 1 (grade 0 vs remaining) and grade $\geq$ 2 (grade 0 and 1 vs remaining). Especially for young patients even relatively mild forms of side effects (grade 1) may have an impact on their QoL, while side effects grade $\geq$ 2 are more severe and therefore clinically more relevant for the general patient population.

All OARs that were contoured during the RT treatment planning process were retrospectively checked and re-delineated. OARs that were not included in the original structure set were retrospectively contoured according to delineation guidelines (Schmahmann et al., 1999; Chera et al., 2009; Sun et al., 2014; Scoccianti et al., 2015; Eekers et al., 2018). All dose distributions were exported using RayStation® scripts. Volume parameters *VxGy*(*RBE*) ranging from 5 Gy(RBE) to 60 Gy(RBE) in increments of 5 Gy(RBE) were analysed. For the non-circumscribed organ skin, absolute volume parameters (in cm^3) were considered. For circumscribed OARs, relative volume parameters (in per cent of the entire volume of the structure) were assessed. Dose parameters *Dx*% ranging from 5 % to 55 % in increments of 10 % as well as *D*2% representing near-maximum

doses were analysed. For OARs with volumes $< 10\,\mathrm{cm}^3$, $D2\%$, D_{mean} and D_{median} ($D50\%$) were considered. For very small structures with volumes $< 2\,\mathrm{cm}^3$ (lacrimal gland and cochlea), only $D2\%$ and D_{mean} were used. For brain stem and optic pathways, the near-maximum parameter $D2\%$ was assessed. For paired OARs, ipsi- and contralateral structures were investigated. For the hippocampus, dose-volume parameters of the bilateral hippocampi were considered. Structures located within the tumour volume were not delineated as related side effects were assumed to be caused rather by the tumour itself than by radiation.

Development of NTCP models

The impact of both dose-volume parameters and clinical variables on the endpoints was evaluated. The clinical parameters comprised age, prescribed total dose, tumour volume (CTV), CTx, tumour location (brain or skull base), gender, and surgery. Uni- and multivariable logistic regression models were developed on the exploration cohort, see equation (2.13), using the *AUC* as a performance metric. The following steps were performed: (I) A 3-fold internal cross-validation was conducted 333 times on the exploration cohort to identify prognostic dose-volume parameters. (II) Dose-volume parameters showing a significant association ($p < 0.05$) to the investigated endpoint in univariable logistic regression and the largest *AUC* value of the internal validation folds were pre-selected. (III) Clinical parameters showing a significant association to the endpoint in logistic regression were tested for correlation with the dose-volume parameters selected during step (II) using the Spearman correlation coefficient ρ. Multivariable logistic regression models containing the independent clinical parameters (in case $|\rho| <$0.5) and one dose-volume parameter were built as described in step (I). Uni- or multivariable models with the largest *AUC* value in internal cross-validation were selected as final NTCP models. (IV) The models derived from the exploration cohort were applied without changes to the validation cohort, i.e. the models were evaluated using the dose-volume parameters of the validation cohort and the model coefficients derived from the exploration cohort. The resulting probabilities were compared to the observed outcome by calculating the *AUC* value. The 2.5^{th} and 97.5^{th} percentiles of 1000 bootstrap samples were used to estimate the 95 % confidence interval (CI) of the *AUC* values for the validation cohort. The prognostic performance was assessed by *AUC* values and calibration plots. Binary variables were compared between the exploration and the validation cohort using exact Fisher tests. Differences in categorical variables were evaluated by χ^2 tests and in continuous variables by Mann-Whitney-U tests. Correlations between dose-volume parameters were assessed using the Spearman correlation coefficient ρ. For all statistical analyses, two-sided tests were performed using IBM SPSS Statistics 25 (IBM Corporation, Armonk, NY) and in-house written Python (version 2.7.10) programmes using the module scipy (statsmodels). P-values < 0.05 were considered statistically significant without adjusting for multiple testing.

For the final models, the regression curves, calibration plots and ROC curves were graphically displayed. The patients were sorted according to the value of the model predictor and subse-

quently divided into nearly equally sized sets. The mean value of each set was represented by one data point. The error bar of each data point indicated the standard deviation of each set.

4.2 Modelling of early side effects

Patient data

The exploration cohort consisted of 113 patients treated within clinical studies (DRKS00007670, 2015; DRKS00008569, 2015; NCT02824731, 2016) at UPTD at the time of analysis (2017). Validation cohort 1 comprised 71 patients treated within a clinical registry study (DRKS00004384, 2012) at the WPE and validation cohort 2 consisted of 96 patients treated at MGH. This retrospective trial was approved by the local Ethics Committee (EK219062016) and the institutional review board at the external institutions. The study design is presented in figure 4.1 and the patient cohorts are characterised in table 4.1.

Early endpoints and associated organs at risk

The treatment-related side effects alopecia, skin erythema, headache, fatigue and nausea were investigated. For alopecia and erythema, the skin including the hair follicle was considered an associated OAR. It was created by cropping the body contour by 3 mm. For the remaining side effects, the brain excluding the target volume (brain-CTV) as well as the brain stem were investigated. The brain included infra- and supratentorial parts of the brain excluding the brain stem.

4.2.1 Results

Glioblastoma was the most common diagnosis in the exploration cohort, whereas meningioma was most frequent in the validation cohorts, leading to significant differences between the cohorts

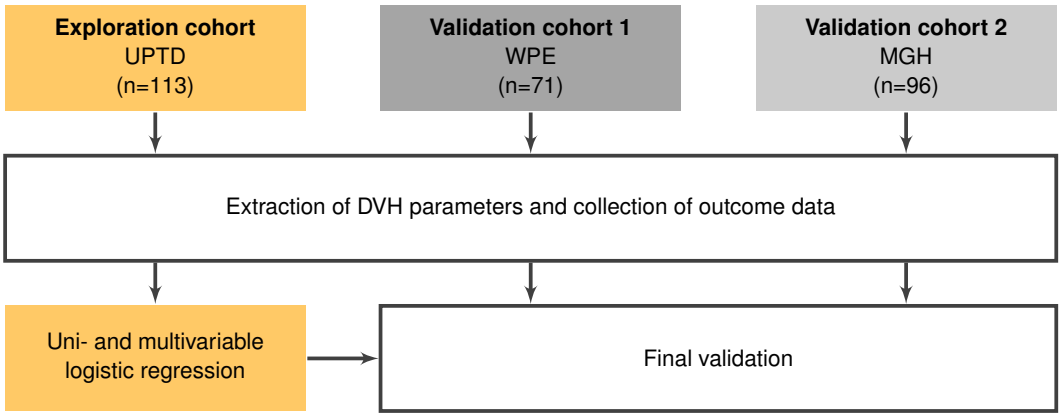


Figure 4.1: Study design to develop and validate models predicting early side effects. UPTD, University Proton Therapy Dresden; WPE, West German Proton Therapy Centre Essen; MGH, Massachusetts General Hospital. Adapted from Dutz et al. (2019).

Table 4.1: Patient characteristics of the exploration cohort and the two validation cohorts for the analysis of early side effects. P-values refer to differences between the exploration and the respective validation cohort.

	Exploration cohort		Validation cohort 1			Validation cohort 2		
Characteristics	**Median**	**(Range)**	**Median**	**(Range)**	**p-value**	**Median**	**(Range)**	**p-value**
Age at PBT in years	49.3	(21.2 – 79.9)	45.1	(18.1 – 79.6)	0.085	45.9	(18.5 – 89.3)	0.094
Tumour volume (CTV) in cm^3	153.9	(1.2 – 499.1)	50.9	(2.1 – 368.1)	< 0.001	28.1	(1.2 – 267.2)	< 0.001
Prescribed dose in Gy(RBE)	60.0	(30.0 – 74.0)	54.0	(42.5 – 60.0)	< 0.001	54.0	(36.0 – 70.2)	< 0.001
	N	**(%)**	**N**	**(%)**	**p-value**	**N**	**(%)**	**p-value**
Gender					0.094			0.003
Male/female	68/45	(60/40)	33/38	(46/54)		38/58	(40/60)	
Surgery					0.011			0.067
Yes/no	99/14	(88/12)	51/20	(72/28)		75/22	(78/22)	
Chemotherapy					0.49			0.45
Yes/no	31/82	(27/73)	16/55	(23/77)		22/74	(23/77)	
Tumour location					0.11			< 0.001
Brain/skull base/other	77/35/1	(68/31/1)	39/30/2	(55/42/3)		37/57/2	(39/59/2)	
Tumour histology*					0.002			< 0.001
High grade glioma	43	(38)	23	(32)		15	(16)	
Low grade glioma	21	(19)	8	(11)		11	(11)	
Meningioma	15	(13)	26	(37)		38	(40)	
Other	34	(30)	14	(20)		32	(33)	

Abbreviations: N, number of patients; CTV, clinical target volume; *Tumour classification according to WHO Classification of Tumours of the Central Nervous System, Fourth Edition.

regarding CTV ($p < 0.001$), prescribed total dose ($p < 0.001$), and surgery ($p = 0.011$ for validation cohort 1). Moreover, there were differences between exploration cohort and validation cohort 2 in terms of gender and tumour location ($p < 0.003$). For all cohorts, the incidence of the investigated acute side effects after PBT is given in table 4.2. Overall, less alopecia and erythema grade 1 – 2 occurred in the validation cohorts compared to the exploration cohort ($p < 0.001$). No statistically significant differences in the frequencies of fatigue and nausea were found between the exploration and validation cohorts. Headaches were less common in validation cohort 2 compared to the exploration cohort ($p < 0.001$). The low incidence rates of nausea grade ≥ 2 in all cohorts did not allow for NTCP modelling.

Side effects were tested for correlation with clinical cofactors in the exploration cohort, see appendix B table B.6. Erythema as well as alopecia occurred significantly more often for larger CTV ($p < 0.024$), prescribed total dose ($p < 0.037$), and performed surgery (erythema grade≥ 1: $p < 0.001$, alopecia grade ≥ 2: $p = 0.030$). Fatigue grade ≥ 1 showed a higher prevalence in women compared to men ($p = 0.005$).

Dose-volume parameters of the skin in the high dose region were significantly correlated to erythema and alopecia on the exploration cohort. For erythema, the parameter revealing the largest *AUC* values in cross-validation was $V35Gy(RBE)$ (grade ≥ 1: $AUC = 0.75$, grade ≥ 2: $AUC = 0.77$). For alopecia grade ≥ 1, $D2\%$ ($AUC = 0.88$) and for alopecia grade ≥ 2, $D5\%$ ($AUC = 0.82$) were selected. As the clinical cofactors associated to erythema and alopecia were significantly correlated with each other and with the selected dose-volume parameters of the skin, see appendix B table B.8, univariable NTCP models including the above-mentioned dose-volume parameters were finally developed, see table 4.3.

Table 4.2: Comparison of the baseline-corrected early side effects (CTCAE v4.0) between the exploration and validation cohorts. The number N of available datasets is given. P-values represent the results of the χ^2 test comparing the exploration cohort with the respective validation cohort.

Side effect		Grade 0		Grade 1		Grade 2		Grade 3		
Cohort	**N**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**p-value**
Alopecia										
Exploration	111	15	(14)	26	(23)	70	(63)			
Validation 1	71	22	(31)	35	(49)	14	(20)			< 0.001
Validation 2	96	59	(61)	16	(17)	21	(22)			< 0.001
Erythema										
Exploration	113	15	(13)	57	(51)	40	(35)	1	(1)	
Validation 1	71	32	(45)	39	(55)	0	(0)	0	(0)	< 0.001
Validation 2	96	63	(66)	32	(33)	1	(1)	0	(0)	< 0.001
Fatigue										
Exploration	112	35	(31)	53	(47)	21	(19)	3	(3)	
Validation 1	71	18	(25)	45	(63)	6	(9)	2	(3)	0.12
Validation 2	96	40	(42)	47	(49)	9	(9)	0	(0)	0.063
Headache										
Exploration	113	57	(51)	33	(29)	18	(16)	5	(4)	
Validation 1	71	49	(69)	15	(21)	6	(8)	1	(2)	0.079
Validation 2	96	75	(78)	19	(20)	2	(2)	0	(0)	< 0.001
Nausea										
Exploration	98	82	(84)	13	(13)	3	(3)	0	(0)	
Validation 1	71	62	(87)	9	(13)	0	(0)	0	(0)	0.32
Validation 2	96	87	(91)	5	(5)	4	(4)	0	(0)	0.15

Abbreviation: n, number of patients.

Table 4.3: Logistic modelling results: Dose-volume parameters of different organs at risk and clinical cofactors for early erythema, alopecia, and fatigue. Mean *AUC* values for 3-fold cross-validation (333 repetitions) and external validation are given. P-values were calculated on the exploration cohort. Fitting parameters β_i as defined in equation (2.13).

Model	β_i	**(95 % CI)**	**p-value**		**AUC**	**(95 % CI)**
Erythema grade $\geq$ 1				Cross validation	0.75	(0.54 – 0.90)
Skin *V*35*Gy*(*RBE*)	0.085	(0.02 – 0.15)	0.008	Validation 1	0.87	(0.77 – 0.95)
Constant	1.00	(0.32 – 1.69)		Validation 2	0.80	(0.71 – 0.89)
Erythema grade $\geq$ 2				Cross validation	0.77	(0.64 – 0.89)
Skin *V*35*Gy*(*RBE*)	0.056	(0.03 – 0.08)	< 0.001	Validation 1	*	
Constant	-1.54	(-2.20 – -0.88)		Validation 2	0.84[†]	(0.77 – 0.91)
Alopecia grade $\geq$ 1				Cross validation	0.88	(0.73 – 0.99)
Skin *D*2%	0.10	(0.05 – 0.15)	< 0.001	Validation 1	0.82	(0.70 – 0.92)
Constant	-0.94	(-2.14 – 0.27)		Validation 2	0.84	(0.75 – 0.92)
Alopecia grade $\geq$ 2				Cross validation	0.82	(0.69 – 0.95)
Skin *D*5%	0.081	(0.05 – 0.11)	< 0.001	Validation 1	0.77	(0.63 – 0.89)
Constant	-1.33	(-2.91 – -0.47)		Validation 2	0.85	(0.76 – 0.92)
Fatigue grade $\geq$ 1				Cross validation	0.68	(0.50 – 0.84)
Brain-CTV *D*2%	0.027	(0.00 – 0.06)	0.067	Validation 1	0.45	(0.31 – 0.61)
Gender	1.28	(0.33 – 2.23)	0.009	Validation 2	0.52	(0.40 – 0.64)
Constant	-0.90	(-2.30 – 0.51)				

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; *External validation not applicable due to zero incidence. [†] Only one event.

The high-dose parameter $D2\%$ of brain-CTV was significantly associated with fatigue grade ≥ 1 and revealed an *AUC* value of 0.60 in exploration. As the cofactor gender was not correlated to this dose-volume parameter ($\rho = 0.15$), a bivariable model was built that showed an improved performance in cross-validation compared to the univariable model (*AUC* = 0.68), see table 4.3. None of the dose-volume parameters of brain-CTV showed significant correlation to fatigue grade ≥ 2, headache, and nausea.

The final NTCP models developed on the exploration cohort were applied without changes to the validation cohorts showing high *AUC* values for erythema and alopecia, see table 4.3. Exemplarily, for erythema grade ≥ 1 and $V35\,Gy(RBE)$ of the skin, the validation *AUC* was 0.87, 95 % CI: [0.77 – 0.95] in validation cohort 1 and 0.80 [0.71 – 0.89] in validation cohort 2. Similar results were obtained for alopecia. Thus, both validation cohorts confirmed the results of the exploration cohort. The regression curves and calibration plots for alopecia grade ≥ 2 and $D5\%$ as well as erythema grade ≥ 1 and $V35\,Gy(RBE)$ of the skin are shown in figure 4.2. The calibration plots for both endpoints showed a right shift for both validation cohorts. Since only one case of erythema grade ≥ 2 was reported in the validation cohorts, the NTCP models for this endpoint could not be validated. External validation of the bivariable NTCP model for fatigue grade ≥ 1 failed with low *AUC* values in both cohorts (validation cohort 1: 0.45 [0.31 – 0.61], validation cohort 2: 0.52 [0.40 – 0.64]).

4.2.2 Discussion

This study on adult patients with intracranial tumours receiving PBT investigated the relation of acute side effects and dose to associated OARs. Overall, PBT was well tolerated with very low incidences of side effects grade 3. Dose-volume parameters in the high-dose region were prognostic for erythema and alopecia. Fatigue was found to be associated with gender and high doses to the brain-CTV. While the NTCP models for erythema as well as alopecia showed similar or even improved performance for the validation cohorts compared to the exploration cohort, the model for fatigue could not be validated.

Several relatively small studies documented acute side effects in the brain tumour patients receiving PBT according to CTCAE v3 or v4 (Hauswald et al., 2012; Weber et al., 2012; Brown et al., 2013; Grosshans et al., 2014; Shih et al., 2015). The incidence rate of alopecia grade 2 in the exploration cohort (63 %) is comparable to the results of Hauswald et al. (2012) (68 %), even though in this study, the 19 glioma grade I-II patients were treated with an active scanning technique and a median dose of 54 Gy(RBE). Wilkinson et al. (2016) assessed 58 patients with grade II glioma treated with PBT doses between 50.4 and 54 Gy(RBE). During treatment, only grade 1 or 2 toxicities were observed with alopecia (81 %), dermatitis (78 %), fatigue (47 %), and headache (40 %) being the most common. These incidence rates of side effects are of the same range as those reported in this thesis. Weber et al. (2012) investigated 39 skull base meningioma patients treated with spot-scanning PBT with a total dose of 56 Gy(RBE) and Shih

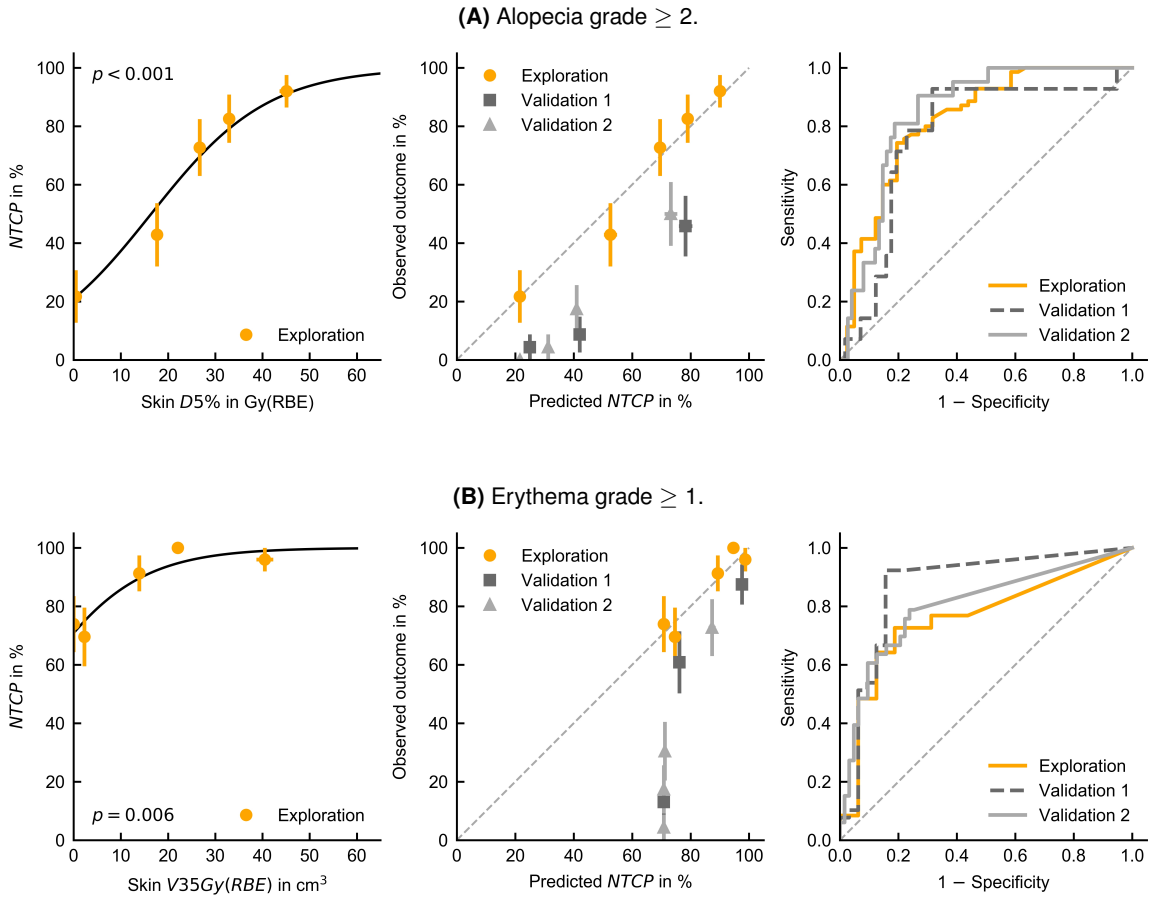


Figure 4.2: NTCP models for acute alopecia grade ≥ 2 and erythema grade ≥ 1. Regression curve (left), calibration plot (centre) and ROC curves (right) are displayed. Each data point and error bar represents the mean value and standard deviation of each patient set, see section 4.1. Adapted from Dutz et al. (2019).

et al. (2015) conducted a study on 20 glioma patients treated with a passive scattering technique with a total dose of 54 Gy(RBE). Both studies did not report any cases of alopecia grade 2, but an incidence rate for alopecia grade 1 of 60 % and 85 %, respectively. In two studies by Grosshans et al. (2014) and Brown et al. (2013) investigating 15 skull base chordoma and chondrosarcoma patients treated with spot-scanning PBT with doses of 69.8 Gy(RBE) and 68.4 Gy(RBE), and 19 medulloblastoma patients receiving a total dose of 54 Gy(RBE), respectively, no erythema grade 2 was documented. This is comparable to validation cohort 1 in this thesis. The incidence rate of fatigue grade ≥ 1 both for the exploration (69 %) and validation cohort (75 %) are in line with the findings of Grosshans et al. (2014) (66 %). For nausea grade ≥ 1, Shih et al. (2015) reported similar results with 20 %, compared to the exploration cohort with 16 %.

The findings of this thesis regarding the association of skin dose-volume parameters with alopecia and erythema are in line with other publications, in which photon doses to the skin

were associated with these side effects. Reduction of the irradiated volume $V24\,Gy$ and $V30\,Gy$ of follicle hair-bearing scalp prevented alopecia in a study using 11-field IMRT (Mahadevan et al., 2015). In this thesis, the skin dose-volume parameter $V25\,Gy(RBE)$ showed also a significant association with alopecia. Nevertheless, the prognostic performance was slightly superior for $D2\%$ (alopecia grade ≥ 1) as well as for $D5\%$ (alopecia grade ≥ 2) in internal cross-validation.

Another study investigated side effects after XRT in 61 cranially irradiated patients (Lawenda et al., 2004). More than 50 % of the patients receiving a follicle dose > 43 Gy developed permanent moderate to severe alopecia. The higher dose value of this prognostic parameter compared to the findings in this thesis may be due to the endpoint of permanent manifestation of alopecia which was investigated in this study. Sung et al. (2016) found the dose-volume parameter $V35\,Gy$ as the most significant predictor for radiation dermatitis grade ≥ 2 after hybrid IMRT for 101 breast cancer patients. As the same dose-volume parameter was found as the most predictive parameter for erythema in this thesis, skin reaction in thoracic regions may be similar to those after cranial irradiation. Mendelsohn et al. (2002) reported early dry desquamation (erythema grade ≥ 1) if the total skin dose during conventional radiotherapy does not exceed 30 Gy and acute moist desquamation (erythema grade ≥ 2 and 3) as an effect of total skin dose > 40 Gy. These findings are in line with the results of this thesis as the reduction of $V35\,Gy(RBE)$ is predicted to lower the risk of erythema. However, in this study, it was observed that patients who had only a low absolute skin volume exposed to 35 Gy(RBE) already had a high probability of erythema grade ≥ 1, see figure 4.2B. This indicates that many patients who had developed erythema were exposed to lower skin doses than 35 Gy(RBE). Thus, a high dose parameter (e.g. $D2\%$ or $D5\%$) that covers a wide range of doses continuously may be more appropriate to describe this dose-effect relationship than a volume parameter which may introduce a threshold effect.

The dose-volume parameters of the skin were highly correlated so that other parameters may be relevant on other cohorts. A principal component analysis was conducted to create independent parameters (principal components) representing different dose regions, see appendix B tables B.9 and B.10. NTCP models based on these components showed similar *AUC* values in internal and external validation, but they are more difficult to interpret. Therefore, they are less suitable for use in clinical routine.

A shift in the calibration curves for alopecia and erythema was observed in both validation cohorts for the presented model as well as for the models based on principal components. Different rates of erythema and alopecia could be caused by differences in several patient and tumour characteristics, such as tumour type and volume, see table 4.1. Nevertheless, other considerable factors such as malnutrition, age, and tobacco abuse history may also be important (Ginot et al., 2010). Different treatment techniques in exploration (DS) and validation cohort 1 (PBS) may also lead to differences in acute side effects. A study investigating skin dose differences between PBS and DS PBT in prostate cancer patients found a lower skin dose in patients receiving actively formed PBT compared to patients treated with passive PBT (Arjomandy et al.,

2009). With cranial irradiation, however, the target is located less deep in the body than with pelvic irradiation. Moreover, in many brain tumour patients, the target volume is located directly under the scalp, so that skin doses differ less between both techniques. A comparison of dose-volume parameters of the skin between the exploration and validation cohort 1 showed significant differences only in the low dose range, see appendix B table B.7. These findings and the shift in calibration for validation cohort 2 (PBS) support the hypothesis, that the treatment techniques are probably not the reason for different toxicity incidences in this thesis. Hence, the differences are most likely due to different toxicity scoring by the physicians, even though the same grading system was used. Therefore, a study investigating the interobserver variability of toxicity assessment was conducted that is described in section 4.4. Di Maio et al. (2015) detected frequent under-reporting of subjective side effects, such as alopecia and fatigue, by physicians at three different centres to a variable extent compared to patients' self-assessment. The incorporation of patient-reported outcome versions of the CTCAE scale (PRO-CTCAE™) in clinical trials may reduce such centre-specific differences in toxicity assessment (Basch et al., 2014).

For alopecia and erythema, univariable NTCP models have been developed. The clinical cofactors surgery, prescribed dose, and CTV were highly correlated to both acute side effects and dose-volume parameters of the skin. As diagnosis determines the dose prescription, tumour volume, and margin size, patients with large CTV had significantly higher prescribed total doses ($p < 0.001$), leading to larger dose-volume parameter values. Due to these high correlations, the clinical cofactors were not included in potential multivariable NTCP models, since including redundant information would not improve their prognostic ability.

For acute fatigue, a bivariable model including the dose-volume parameter $D2\%$ of brain-CTV and the clinical cofactor gender was developed in this thesis. This model could not be validated. Gulliford et al. (2012a) investigated the association of dose to cranial structures and acute fatigue in 67 patients treated with 3D-CRT or IMRT and found significantly higher mean and maximum dose values for posterior fossa, brain stem, and cerebellum for patients suffering from acute fatigue grade ≥ 2 ($p < 0.01$). Ferris et al. (2018) found a significant correlation between patient-reported fatigue based on Multidimensional Fatigue Inventory scores and maximum dose to the brain stem as well as medulla in 124 patients treated with IMRT or VMAT ($p < 0.05$). In this thesis, no statistically significant dose relationship between dose to the brain stem and acute fatigue for patients treated with passively scattered proton beams could be observed; instead, a significant correlation between fatigue grade ≥ 1 and female gender was found. One of the few studies investigating cancer-related fatigue according to the EORTC QLQ-C30 scale also found female gender associated with greater fatigue severity (Pater et al., 1997). However, as there are only a few studies on gender differences in cancer-related fatigue, there is still concern about the prevalence and severity of this side effect (Miaskowski, 2004).

Finally, no association between dose to the brain and acute nausea or headache was found in this study. These endpoints are difficult to measure and may also be alternating. Radiation dose

to brain structures or concomitant medication such as analgesics may play a role. No NTCP models based on XRT or PBT data regarding these endpoints are available in the literature.

4.3 Modelling of late side effects

Patient data

Patient data from the three cohorts presented in chapter 3 were partly used to model late side effects. Patients treated at UPTD were enrolled in two clinical registry trials (DRKS00007670, 2015; DRKS00008569, 2015). Long-term follow-up data were not available for all patients. Some patients did not attend continuous visits, e.g. because they lived at great distances from the proton therapy centres. Others developed a tumour recurrence or died. Thus, the patient cohorts from UPTD and WPE were combined to a single exploration cohort to increase the number of patients. The validation cohort consisted of patients treated at MGH. In contrast to the analysis of acute side effects, the follow-up intervals differed markedly between the three cohorts. While follow-up data were available every three months for patients treated at UPTD, data were only available every six months for patients treated at MGH and only every 12 months for patients treated WPE. It was decided not to evaluate the maximum value of severity grades (e.g. during one year), as this may have introduced bias due to the different follow-up periods. In patients who are seen four times a year, the probability of diagnosing a more severe side effect may be increased compared to patients who are seen once a year, especially for side effects that may alternate or due to interobserver variability. Hence, to ensure a comparable evaluation between the centres, the models were developed for side effects occurring at 12 months and at 24 months following PBT, see figure 4.3. Thus, at 12 months following PBT, 104 and 67 patients were included in the exploration and validation cohort, respectively. Similarly, at 24 months following PBT, 74 and 65 patients were included in the exploration and validation cohort, respectively. The patient cohorts are characterised in table 4.4 for the exploration and validation cohorts at 12 months and at 24 months after PBT. Due to the low incidence rates, additional models were developed using the pooled cohorts at 12 months and at 24 months after PBT, see figure 4.3.

To consider all available patient data from all follow-up visits between 6 and 24 months following PBT, GEEs on the pooled cohort including 216 patients were set up as described below (see also section 2.4). The characteristics of the pooled cohort used in GEE analysis is presented in appendix B table B.11.

This retrospective trial was approved by the local Ethics Committee (EK566122019) and the institutional review board at the external institutions.

Late endpoints and associated organs at risk

Different physician-rated late side effects were recorded at the three centres. In all institutions, the following side effects were scored according to CTCAE: alopecia, cataract, dry eye syndrome,

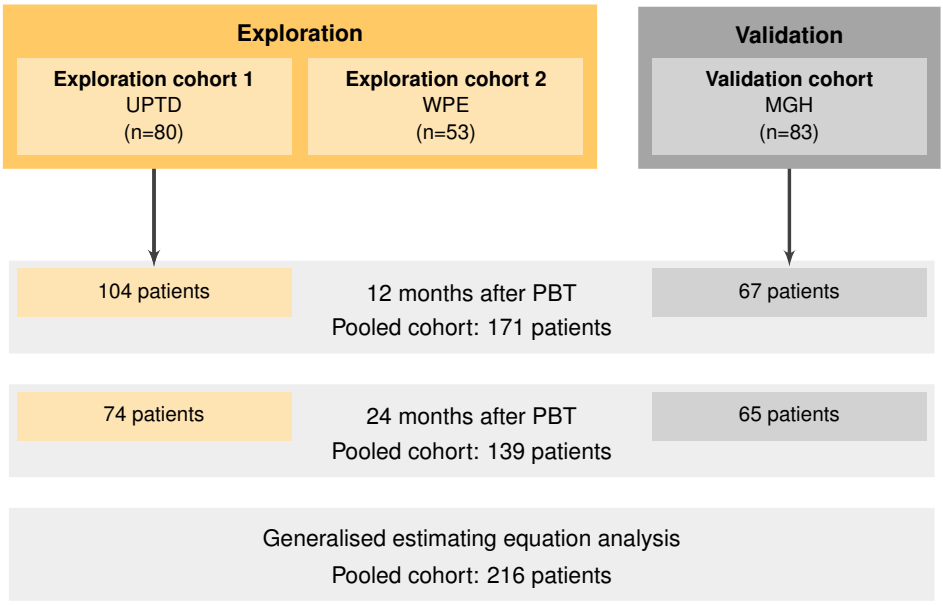


Figure 4.3: Study design to develop and validate late side effects. UPTD, University Proton Therapy Dresden; WPE, West German Proton Therapy Centre Essen; MGH, Massachusetts General Hospital.

Table 4.4: Patient characteristics of the exploration cohort and the validation cohort for the analysis of late side effects at 12 months and at 24 months after PBT.

	12 months following PBT					24 months following PBT				
	Exploration cohort		Validation cohort			Exploration cohort		Validation cohort		
Total number of patients	104		67			74		65		
Characteristics	**Median**	**(Range)**	**Median**	**(Range)**	**p-value**	**Median**	**(Range)**	**Median**	**(Range)**	**p-value**
Age[1] at PBT in years	48.2	(18.1 – 84.7)	45.9	(21.6 – 89.3)	0.22	45.3	(18.1 – 78.0)	47.0	(21.9 – 78.2)	0.93
Tumour volume (CTV)[1] in cm^3	71.3	(1.2 – 498.0)	29.6	(1.2 – 267.2)	0.003	61.2	(1.2 – 370.7)	29.4	(2.7 – 267.2)	0.015
Total dose[1] in Gy(RBE)	54.0	(30.0 – 74.0)	54.0	(45.0 – 70.2)	< 0.001	54.0	(30.0 – 74.0)	54.0	(45.0 – 70.2)	< 0.001
	N	**(%)**	**N**	**(%)**	**p-value**	**N**	**(%)**	**N**	**(%)**	**p-value**
Gender[2]										
Male/female	55/49	(53/47)	26/41	(39/61)	0.085	38/36	(51/49)	27/38	(42/58)	0.31
Surgery[2]										
No/yes/missing	22/81/1	(21/78/1)	12/55/0	(18/82/0)	0.70	14/60	(19/81)	12/53	(18/82)	1.00
Chemotherapy[2]										
No/yes/missing	76/27/1	(73/26/1)	48/19	(72/28)	0.86	57/17	(77/23)	49/16	(75/25)	0.84
Re-irradiation[2]										
No/yes/missing	95/7/2	(91/7/2)	64/3	(96/4)	0.74	69/4/1	(93/5/1)	64/1/0	(98/2/0)	0.37
Tumour location[2]										
Brain/skull base/other	49/53/2	(47/51/2)	32/34/1	(48/51/1)	0.98	36/37/1	(49/50/1)	29/35/1	(39/47/1)	0.89
Tumour location[2]										
Left hemisphere	32	(31)	24	(36)	0.42	21	(28)	24	(37)	0.57
Right hemisphere	44	(42)	20	(30)		31	(42)	20	(31)	
Central	25	(24)	20	(30)		19	(26)	18	(28)	
Bilateral	3	(3)	3	(4)		3	(4)	3	(5)	
Tumour location[2]										
Temporal lobe	33	(32)	20	(30)	0.003	24	(32)	21	(32)	0.002
Frontal lobe	25	(24)	15	(22)		20	(27)	15	(23)	
Parietal lobe	2	(2)	7	(10)		1	(1)	6	(9)	
Occipital lobe	3	(3)	1	(1)		3	(4)	0	(0)	
Multiple lobes	23	(22)	3	(4)		16	(22)	3	(5)	
Other	18	(17)	21	(31)		10	(14)	20	(31)	
Tumour histology[2]										
High-grade glioma	25	(24)	13	(19)	0.68	16	(22)	11	(17)	0.63
Low-grade glioma	16	(15)	11	(16)		11	(15)	11	(17)	
Meningioma	29	(28)	24	(36)		23	(31)	26	(40)	
Other	34	(33)	19	(28)		24	(32)	17	(26)	

Abbreviations: N, number of patients; CTV, clinical target volume; [1]Mann-Whitney U test, [2] χ^2 test.

fatigue, headache, hearing and memory impairment, optic nerve failure, and seizure. While some of these side effects were too unspecific to correlate them to dose-volume parameters of specific OARs (e.g. seizure), the incidence rates for other side effects have been too low for developing NTCP models (e.g. cataract). The side effects that were finally considered for NTCP modelling and the potentially associated OARs are given in table 4.5. They were dichotomised and analysed separately for grade ≥ 1 and ≥ 2, if at least 8 incidences occurred, see table 4.6.

Experimental design

The analyses of late side effects followed the same methodology as described in section 4.1. A 3-fold internal cross-validation was conducted 333 times on the exploration cohort to identify prognostic dose-volume parameters. Dose-volume parameters showing a significant association ($p < 0.05$) to the investigated endpoint in univariable logistic regression and the largest *AUC* value of the internal validation folds were pre-selected. Clinical parameters showing a significant association to the endpoint in logistic regression were tested for correlation with the selected dose-volume parameters. Multivariable logistic regression models containing the independent clinical parameters and one dose-volume parameter were built. Uni- or multivariable models with the largest *AUC* value in internal cross-validation were selected as final NTCP models. However, only one instead of two validation cohorts was used.

Since the incidence rates of various side effects were low, an additional exploratory analysis on the pooled data of all cohorts was conducted without external validation at 12 months and at 24 months after PBT. For this pooled cohort, univariable logistic regression using dosimetric parameters (see table 4.5) and clinical cofactors was applied. The clinical parameters comprised age at PBT, prescribed total dose, tumour volume (CTV), CTx, tumour location (brain or skull base), gender, and surgery. The dosimetric parameter of the model with the highest *AUC* value among all models including dosimetric parameters was determined. Subsequently, this dosimetric parameter as well as all clinical cofactors that were significantly associated with the endpoint

Table 4.5: Investigated late side effects, associated OARs and the investigated dosimetric parameters.

Side effect	Associated organs at risk	Dosimetric paremeters
Alopecia	Skin	$D2\%, D5\%, D15\%\ldots, D55\%, V5Gy(RBE), V10Gy(RBE), \ldots, V60Gy(RBE)$ in cm^3
Dry eye syndrome	Lacrimal gland (ipsilateral)	$D2\%, D_{mean}$
Fatigue	Brain-CTV, cerebellum Brain stem	$D2\%, D50\%, D_{mean}, V5Gy(RBE), V10Gy(RBE), \ldots, V60Gy(RBE)$ in % of entire volume $D2\%$
Headache	Brain-CTV, cerebellum Brain stem	$D2\%, D50\%, D_{mean}, V5Gy(RBE), V10Gy(RBE), \ldots, V60Gy(RBE)$ in % of entire volume $D2\%$
Hearing impairment	Cochlea (ipsilateral)	$D2\%, D_{mean}$
Memory impairment	Brain-CTV, cerebellum Hippocampi bilateral	$D2\%, D50\%, D_{mean}, V5Gy(RBE), V10Gy(RBE), \ldots, V60Gy(RBE)$ in % of entire volume $D2\%, D50\%, D_{mean}$
Optic nerve failure	Optic nerve (ipsilateral), chiasm	$D2\%$

were included in multivariable models using stepwise forward variable selection based on likelihood ratio using IBM SPSS Statistics 25. The inclusion testing was based on the significance of the score statistic (p = 0.05), and exclusion testing (p = 0.10) based on the probability of a likelihood-ratio statistic based on the maximum partial likelihood estimates. The 95 % confidence intervals of the *AUC* values are asymptotic estimates.

To consider all recorded patient data from all follow-up visits, GEEs were set up to assess which dose-volume parameters were significantly associated with the dichotomised side effects. GEEs allow for the analysis of longitudinal data and account for the non-linear nature of the binary response as well as missing observations. A logit link function and an independent working correlation matrix was used. For each endpoint and dosimetric or clinical parameter, two GEEs were created. One included the dosimetric or clinical parameter and time as model predictors. The other model additionally included the interaction between time and the dosimetric or clinical parameter. The parameter time was defined as natural numbers in the interval [1, 7] representing the follow-up visits, for equidistant (three months) intervals starting with 1 at 6 months after PBT and ending with 7 at 24 months after PBT. For each side effect and grade, the model was selected, which revealed the lowest quasi log-likelihood estimate and the corresponding dosimetric or clinical predictor showed the lowest p-value. Additionally, the predictor was preferred, that showed a significant association with the endpoint in both models including and excluding the interaction term. These analyses were performed for the above-mentioned side effects on the pooled cohort using IBM SPSS Statistics 25.

4.3.1 Results

Development of NTCP models and their external validation

Incidence rates of late side effects The exploration and validation cohort differed in a few clinical characteristics at both time points of analysis (at 12 and at 24 months after PBT), see table 4.4. The exploration cohort included patients with significantly larger tumours than the validation cohort (median: 71.3 vs 29.6 cm^3). The prescribed dose was higher in the exploration cohort than in the validation cohort. Moreover, the cohorts differed significantly with respect to the location of the tumours. In the exploration cohort, more patients had a tumour involving more than one lobe compared to the validation cohort.

The incidence rates for late side effects at 12 and at 24 months following PBT are given in table 4.6. Overall, the incidence rates of late radiation-induced side effects were low. At 12 months after PBT, fatigue was the most common side effect. Fatigue grade $\geq$ 1 was observed in 36 % and 26 % of the patients in the exploration and validation cohort, respectively. Fatigue grade $\geq$ 2 was observed in 15 % and 13 %, mild alopecia (grade $\geq$ 1) in 34 % and 22 % and headache grade $\geq$ 1 in 25 % and 12 % of the patients in the exploration and validation cohort, respectively. In three patients in the exploration cohort, optic nerve failure grade 4 was observed. The incidence rates of the side effects did not differ between the exploration and validation cohort,

Table 4.6: Comparison of the baseline-corrected late side effects (CTCAE v4.0) between exploration and validation cohort at 12 and 24 months after PBT. The number N of available datasets is given. P-values represent results of the χ^2 test. Side effects considered for modelling are coloured.

	12 months after PBT												24 months after PBT											
Toxicity		**Grade 0**		**Grade 1**		**Grade 2**		**Grade 3**		**Grade 4**				**Grade 0**		**Grade 1**		**Grade 2**		**Grade 3**		**Grade 4**		
Cohort	**N**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**p-value**	**N**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**n**	**(%)**	**p-value**
Alopecia												0.26												0.057
Exploration	84	55	(65)	27	(32)	2	(2)						65	46	(71)	18	(28)	1	(2)					
Validation	45	35	(78)	10	(22)								35	32	(91)	3	(9)							
Dry eye syndrome												0.034												0.42
Exploration	102	92	(90)	6	(6)	4	(4)						70	68	(97)	1	(1)	1	(1)					
Validation	65	65	(100)										59	59	(100)									
Fatigue												0.56												0.004
Exploration	104	67	(64)	22	(21)	13	(13)	2	(2)				73	45	(62)	21	(29)	7	(10)					
Validation	62	46	(74)	8	(13)	7	(11)	1	(2)				60	52	(87)	7	(12)	1	(2)					
Headache												0.12												0.064
Exploration	104	78	(75)	16	(15)	7	(7)	3	(3)				73	57	(78)	9	(12)	7	(10)					
Validation	62	55	(89)	6	(10)	1	(2)						61	52	(85)	8	(13)	0	(0)	1	(2)			
Hearing impairment												0.72												0.21
Exploration	102	93	(91)	4	(4)	2	(2)	3	(3)				70	62	(89)	5	(7)	2	(3)	0	(0)	1	(1)	
Validation	64	59	(92)	3	(5)	0	(0)	2	(3)				59	55	(93)	2	(3)	0	(0)	2	(3)			
Memory impairment												0.061												0.011
Exploration	103	78	(76)	16	(16)	8	(8)	1	(1)				73	50	(68)	17	(23)	6	(8)					
Validation	63	58	(92)	4	(6)	1	(2)						60	53	(88)	7	(12)	0	(0)					
Optic nerve failure												0.10												0.17
Exploration	100	86	(86)	7	(7)	3	(3)	1	(1)	3	(3)		71	59	(83)	6	(8)	2	(3)	2	(3)	2	(3)	
Validation	62	61	(98)	0	(0)	1	(2)						56	54	(96)	1	(2)	1	(2)					
Cataract												0.94												0.36
Exploration	57	56	(98)	1	(2)								36	32	(89)	2	(6)	1	(3)	1	(3)			
Validation	66	66	(100)										62	60	(97)	1	(2)	0	(0)	1	(2)			
Seizure												0.53												0.18
Exploration	101	97	(96)	0	(0)	3	(3)	1	(1)				70	67	(96)	0	(0)	2	(3)	1	(1)			
Validation	64	61	(95)	1	(2)	2	(3)						61	59	(97)	2	(3)							

Abbreviation: n, number of patients.

except for dry eye syndrome (p = 0.034). Dry eye syndrome grade $\geq$ 1 was observed in 10 % of the patients in the exploration cohort, but not in the validation cohort.

At 24 months following PBT, fatigue was still the most frequent side effect. Fatigue grade $\geq$ 1 was reported by 39 % and 16 %, memory impairment grade $\geq$ 1 in 31 % and 12 %, mild alopecia (grade $\geq$ 1) occurred in 30 % and 9 % of the patients in the exploration and validation cohort, respectively. In one and two patients of the exploration cohort, hearing impairment and optic nerve failure grade 4 was observed, respectively. The incidence rates of the side effects differed between the exploration and validation cohort for fatigue (p = 0.004) and memory impairment (p = 0.011). Both side effects occurred with higher incidence rates and severity grades in the exploration cohort.

The low incidence rates in the exploration cohort did not allow for NTCP modelling of several side effects. At 12 months following PBT, the following side effects were considered: alopecia, dry eye syndrome, headache, optic nerve failure, memory and hearing impairment grade $\geq$ 1 as well as fatigue, headache and memory impairment grade $\geq$ 1 and $\geq$ 2. At 24 months after PBT, alopecia, fatigue, headache, hearing and memory impairment and optic nerve failure grade $\geq$ 1 were investigated.

NTCP models for side effects at 12 months after PBT Alopecia grade ≥ 1 was significantly associated to a wide range of dosimetric parameters of the skin. In cross-validation, the highest *AUC* value (0.93) was observed for two models, one including $V45Gy(RBE)$ of the skin as model predictor and the other model included the near-maximum dose *D*2% of the skin. Both parameters $V45Gy(RBE)$ and *D*2% were also highly correlated (Spearman ρ = 0.95). Alopecia was also associated to tumour volume (CTV), prescribed total dose, surgery, and location in the skull base or brain ($p < 0.02$). In bivariable models, including one of the before-mentioned dosimetric parameters and one of the significant clinical cofactors, the highest *AUC* value was observed for a model including *D*2% of the skin and tumour volume (*AUC* = 0.93 [0.75 – 1.00]). As both parameters were correlated (Spearman ρ = 0.75) and the *AUC* value did not further increase, a univariable model was preferred. For the model including $V45Gy(RBE)$ of the skin and the tumour volume, the *AUC* value even decreased (0.91 [0.74 – 1.00]), while the model including tumour volume as the only parameter had a smaller *AUC* value (*AUC* = 0.83 [0.69 – 0.95]). In external validation, the model including *D*2% performed slightly better (*AUC* = 0.77 [0.64 – 0.89]) and showed better calibration than the model including $V45Gy(RBE)$ (*AUC* = 0.65 [0.47 – 0.83]), see figure 4.4. Characteristics of both models are summarised in table 4.7.

Mild physician-rated memory impairment grade ≥ 1 was significantly associated to the prescribed total dose as the only predictive factor (p = 0.038, *AUC* = 0.64 [0.53 – 0.75]). However, this model could not be validated externally (*AUC* = 0.45 [0.19 – 0.69]). An association with dose-volume parameters of the considered brain structures could not be found.

Hearing impairment grade ≥ 1 was significantly associated with age at PBT (p = 0.002, *AUC* = 0.83 [0.53 – 1.00]). In external validation, this model achieved *AUC* = 0.73 [0.40 – 0.96]. However, the 95 % CI included the value 0.5, which represents a random prediction. Moreover, no dose-effect relationship could be observed.

For headache, fatigue, dry eye syndrome and optic nerve failure grade ≥ 1 as well as fatigue, headache and memory impairment grade ≥ 2, no significant association with either dose-volume parameters or clinical cofactors could be found in logistic regression.

NTCP models for side effects at 24 months after PBT Alopecia grade ≥ 1 was significantly associated to a wide range of dose-volume parameters of the skin. The model including $V30Gy(RBE)$ of the skin showed the highest *AUC* value in cross-validation (0.80 [0.62 – 0.95], p = 0.005). The model including *D*2% of the skin showed a similar *AUC* value and revealed the lowest p-value (*AUC* = 0.78 [0.60 – 0.93], $p < 0.001$). Alopecia was also associated to tumour volume (CTV), prescribed total dose, CTx, and location in the skull base or brain ($p < 0.04$). In bivariable models, including one of the before-mentioned dosimetric parameters and one of the significant clinical cofactors, the *AUC* value did not further increase. The dosimetric parameters $V30Gy(RBE)$ and *D*2% were highly correlated (Spearman ρ = 0.81). Both univariable models including $V30Gy(RBE)$ as well as *D*2% revealed high *AUC* values in external validation (0.90 [0.71 – 1.00] and 0.88 [0.69 – 1.00], respectively). However, external validation was challenging

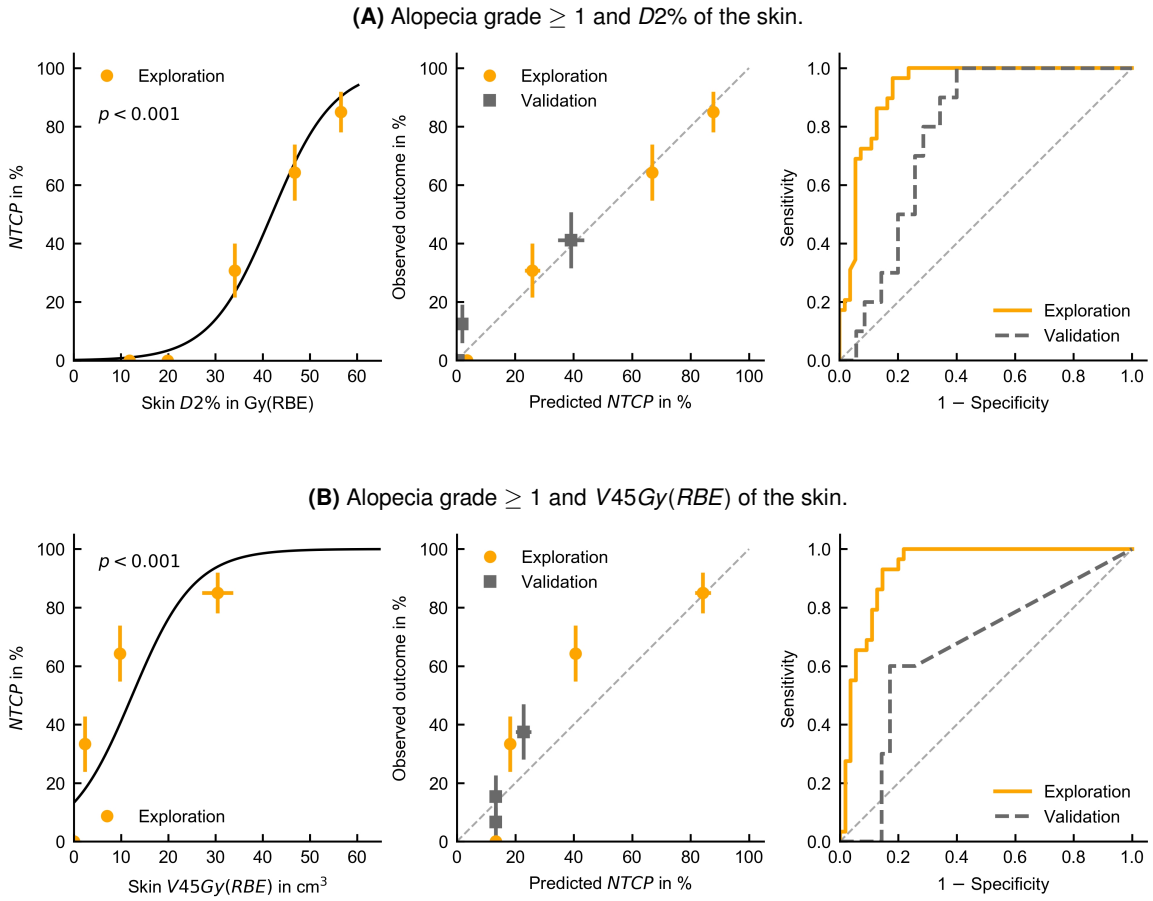


Figure 4.4: NTCP models for alopecia grade $\geq$ 1 at 12 months following PBT with the model parameters *D*2% **(A)** and *V*45*Gy*(*RBE*) of the skin **(B)**. Regression curve (left), calibration plot (centre) and ROC curves (right) are displayed. Each data point and error bar represents the mean value and standard deviation of each patient set, see section 4.1.

due to the low number of incidences of this side effect in the validation cohort, see table 4.6. This is also reflected in the non-ideal calibration, see figure 4.5. The model characteristics are presented in table 4.7.

Hearing impairment grade $\geq$ 1 was significantly associated with mean and near-maximum dose to the ipsilateral cochlea ($p < 0.041$). The model including the mean cochlea dose revealed the highest *AUC* value in cross-validation ($p = 0.035$, $AUC = 0.74$ [0.42 – 1.00]). The high validation *AUC* value of 0.87 [0.62 – 1.00] in the external cohort may be due to the low number of events in this cohort. The low event rate of mild hearing impairment did not allow for multivariable modelling. The model is presented in figure 4.6 and table 4.7. No significant association to age at treatment could be observed ($p = 0.20$)

For memory impairment, headache and optic nerve failure grade $\geq$ 1, no significant association with dose-volume parameters could be found in logistic regression.

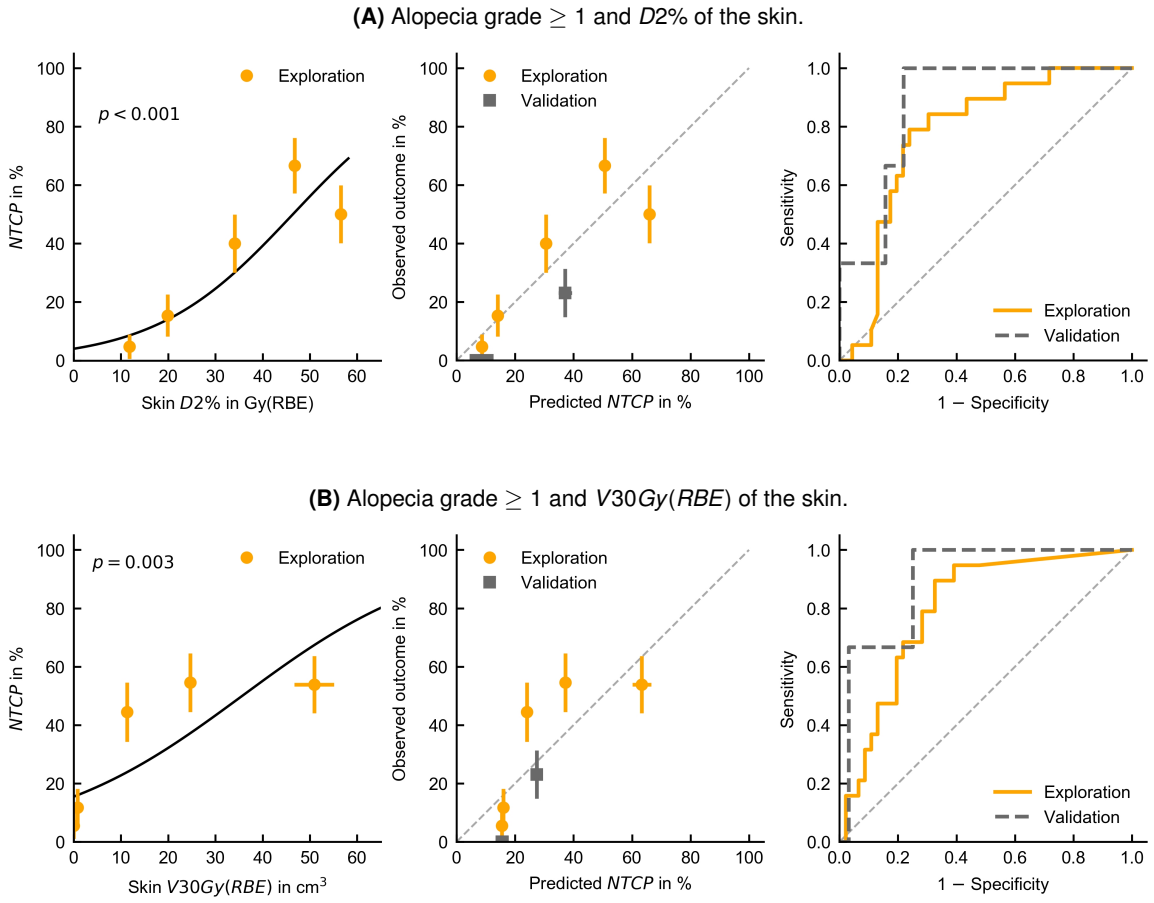


Figure 4.5: NTCP models for alopecia grade $\geq$ 1 at 24 months following PBT with the model parameters *D*2% **(A)** and *V*30*Gy*(*RBE*) of the skin **(B)**. Regression curve (left), calibration plot (centre) and ROC curves (right) are displayed. Each data point and error bar represents the mean value and standard deviation of each patient set, see section 4.1.

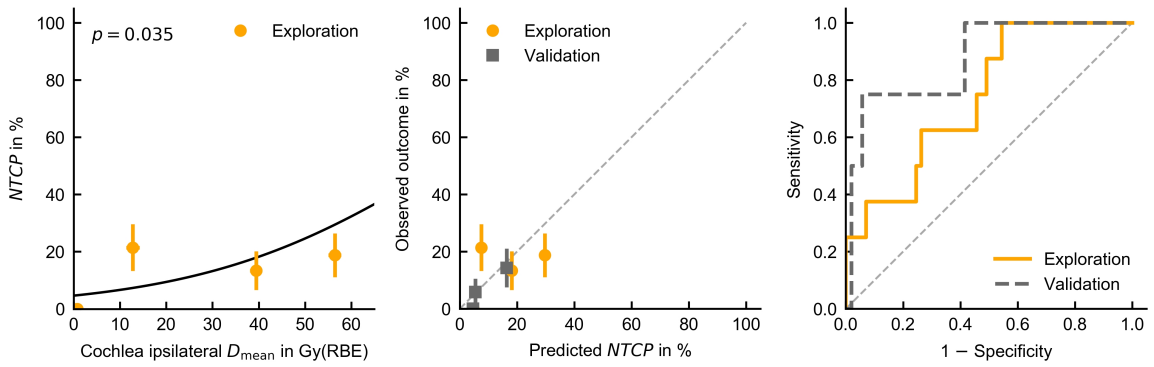


Figure 4.6: NTCP model for hearing impairment grade $\geq$ 1 at 24 months following PBT. Regression curves (left), calibration plot (centre) and ROC curve (right) are displayed. Each data point and error bar represents the mean value and standard deviation of each patient set, see section 4.1.

Table 4.7: NTCP models for late side effects at 12 months and at 24 months following PBT. Model coefficients β_i as described in equation 2.13.

Model	β_i	(95 % CI)	p-value		AUC	(95 % CI)
12 months after PBT						
Alopecia grade ≥ 1						
Skin $V45Gy(RBE)$ in cm^{-3}	0.15	(0.08 – 0.22)	< 0.001	Exploration	0.93	(0.85 – 1.00)
Constant	-1.88	(-2.63 – -1.13)		Validation	0.65	(0.47 – 0.83)
Skin $D2\%$ in $Gy(RBE)^{-1}$	0.15	(0.09 – 0.21)	< 0.001	Exploration	0.93	(0.84 – 0.99)
Constant	-6.38	(-9.15 – -3.60)		Validation	0.77	(0.64 – 0.89)
24 months after PBT						
Alopecia grade ≥ 1						
Skin $V30Gy(RBE)$ in cm^{-3}	0.048	(0.02 – 0.08)	0.003	Exploration	0.80	(0.62 – 0.95)
Constant	-1.70	(-2.52 – -0.88)		Validation	0.90	(0.71 – 1.00)
Skin $D2\%$ in $Gy(RBE)^{-1}$	0.068	(0.03 – 0.11)	0.001	Exploration	0.77	(0.60 – 0.93)
Constant	-3.18	(-4.74 – -1.62)		Validation	0.88	(0.69 – 1.00)
Hearing impairment grade ≥ 1						
Cochlea ipsilateral D_{mean} in $Gy(RBE)^{-1}$	0.038	(0.00 – 0.07)	0.035	Exploration	0.74	(0.42 – 1.00)
Constant	-3.03	(-4.47 – -1.58)		Validation	0.87	(0.62 – 1.00)

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval.

Development of NTCP models based on pooled data

Due to the low incidence rates of several side effects, an exploratory analysis on the pooled cohort, combining the exploration and validation cohort, was conducted. Side effects with at least 10 incidences were considered. At 12 months after PBT, these comprised alopecia, dry eye syndrome, fatigue, headache, hearing and memory impairment as well as optic nerve failure grade ≥ 1 as well as fatigue, headache and memory impairment grade ≥ 2. At 24 months, the side effects alopecia, fatigue, headache, hearing and memory impairment as well as optic nerve failure grade ≥ 1 were investigated.

NTCP models for side effects at 12 months after PBT Alopecia grade ≥ 1 was significantly associated to several dosimetric parameters of the skin ($D2\%$, $D5\%$, $D15\%$, $D25\%$, $V10Gy(RBE)$, ..., $V50Gy(RBE)$, $p < 0.02$). The near-maximum dose parameter $D2\%$ of the skin revealed the highest *AUC* value (0.90 [0.85 – 0.95], $p < 0.001$) similar to the exploration-validation study described before. This side effect was also associated to the clinical cofactors surgery, tumour volume (CTV) and prescribed dose ($p < 0.010$). In multivariable modelling including $D2\%$ and these clinical cofactors, the model predictors $D2\%$ of the skin and prescribed dose remained after forward variable selection. Both parameters were correlated (Spearman $\rho = 0.69$). Moreover, the *AUC* value of this bivariable model did not increase substantially compared to the univariable model including only $D2\%$ of the skin (0.91 [0.81 – 0.99]). Thus, the univariable model including $D2\%$ was preferred as the *AUC* value of the model including prescribed dose only was lower (0.81 [0.73 – 0.89]). Model characteristics are given in table 4.8.

Table 4.8: NTCP models for late side effects in the pooled analysis. The 95 % confidence intervals of the *AUC* values are asymptotic estimates.

Model	β_i	(95 % CI)	p-value	AUC	(95 % CI)
12 months after PBT					
Alopecia grade ≥ 1					
Skin *D*2% in Gy(RBE)$^{-1}$	0.12	(0.08 – 0.15)	< 0.001	0.90	(0.85 – 0.95)
Constant	-4.79	(-6.35 – -3.24)			
Memory impairment grade ≥ 1					
Hippocampi *D*2% in Gy(RBE)$^{-1}$	0.023	(0.004 – 0.043)	0.017	0.66	(0.55 – 0.77)
Constant	-2.32	(-3.12 – -1.51)			
Memory impairment grade ≥ 2					
Brain-CTV $V25Gy(RBE)^*$	5.02	(0.03 – 10.01)	0.049	0.70	(0.52 – 0.88)
Constant	-3.42	(-4.47 – -2.38)			
Hearing impairment grade ≥ 1					
Cochlea ipsi D_{mean} in Gy(RBE)$^{-1}$	0.032	(0.00 – 0.06)	0.021	0.82	(0.70 – 0.93)
Age in years	0.072	(0.03 – 0.12)	0.001		
Constant	-7.02	(-9.86 – -4.18)			
24 months after PBT					
Alopecia grade ≥ 1					
Skin *D*2% in Gy(RBE)$^{-1}$	0.078	(0.04 – 0.11)	< 0.001	0.82	(0.75 – 0.89)
Constant	-3.80	(-5.26 – -2.34)			
Memory impairment grade ≥ 1					
Brain-CTV $V35Gy(RBE)^*$	6.50	(1.30 – 11.70)	0.014	0.64	(0.52 – 0.75)
Constant	-1.77	(-2.40 – -1.15)			
Fatigue grade ≥ 1					
Brain stem *D*2% in Gy(RBE)$^{-1}$	0.021	(0.00 – 0.04)	0.035	0.68	(0.58 – 0.77)
Chemotherapy	-1.16	(-2.30 – -0.02)	0.047		
Constant	-1.52	(-2.39 – -0.64)			
Hearing impairment grade ≥ 1					
Cochlea ipsi D_{mean} in Gy(RBE)$^{-1}$	0.050	(0.019 – 0.081)	0.001	0.79	(0.68 – 0.90)
Constant	-3.48	(-4.69 – -2.28)			

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; *as fraction of the total volume.

Hearing impairment grade ≥ 1 was significantly associated with D_{mean} (p = 0.007) and *D*2% of the ipsilateral cochlea (p = 0.010) in logistic regression. The model including D_{mean} revealed the higher *AUC* value (0.70 [0.55 – 0.85]). Moreover, a significant relation between the clinical cofactor age at treatment and hearing impairment was observed ($p < 0.001$, *AUC* = 0.79 [0.66 – 0.92]). The bivariable model including D_{mean} and age at PBT revealed the highest *AUC* value (0.82 [0.70 – 0.93]). As both parameters were not correlated (ρ = 0.09), this model was preferred, see figure 4.7A and table 4.8.

Memory impairment grade ≥ 1 was significantly associated to D_{mean} and *D*2% of the bilateral hippocampi (p = 0.033 and p = 0.017), $V20Gy(RBE)$ of the brain-CTV (p = 0.040), *D*2%, $V20Gy(RBE)$ and $V20Gy(RBE)$ of the cerebellum ($p < 0.043$). The model with the highest *AUC* value included the dosimetric parameter *D*2% of the hippocampi (0.66 [0.55 – 0.77]). The clinical cofactor prescribed dose was also related to memory impairment grade ≥ 1 (p = 0.017). However, during logistic regression including *D*2% of the hippocampi and prescribed dose using

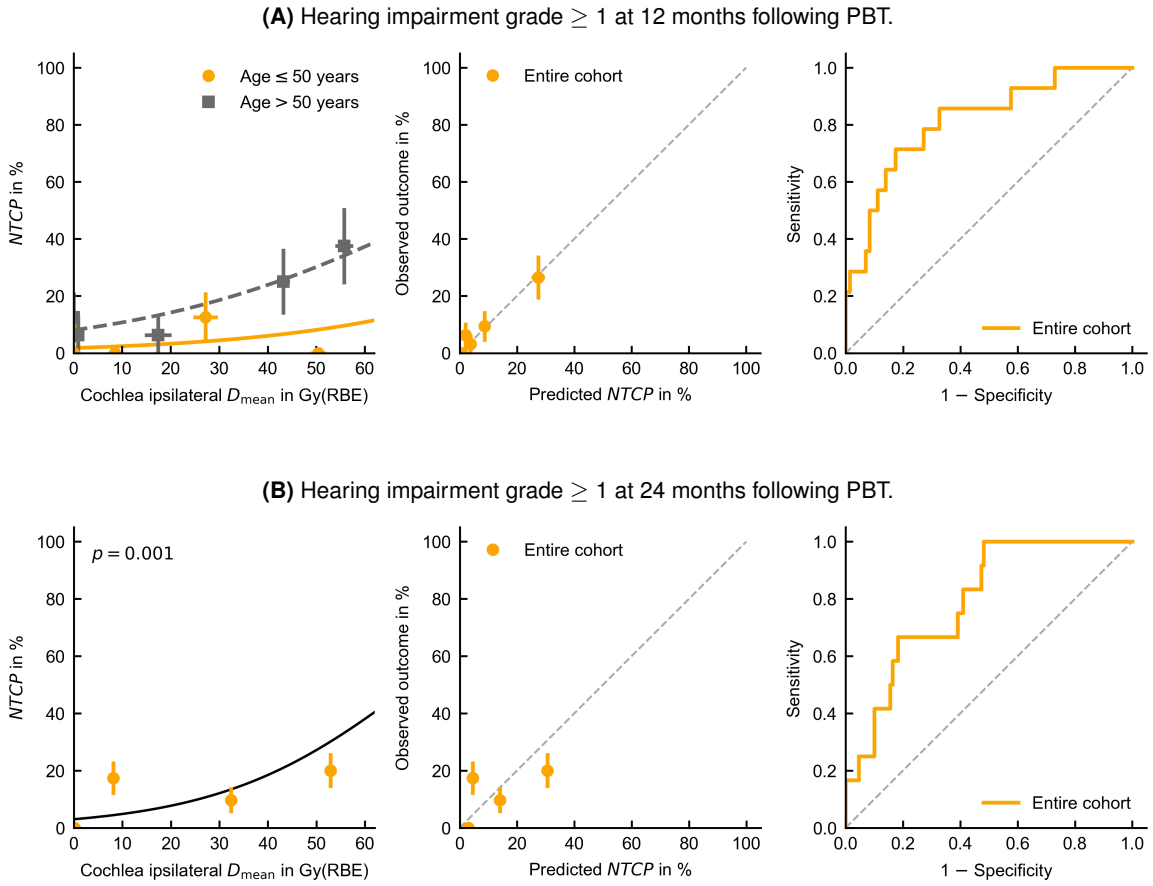


Figure 4.7: NTCP models for hearing impairment grade $\geq$ 1 at 12 months (**A**) and 24 months (**B**) following PBT in the pooled analysis. Regression curve (left), calibration plot (centre) and ROC curve (right) are displayed. The data points represent nearly equally sized patient groups, sorted according to the value of the dosimetric parameter. Each data point and error bar represents the mean value and standard deviation of each patient set, see section 4.1. For the model at 12 months after PBT, two regression curves for two different age groups are shown as an illustration. The calibration plot and ROC curve refer to the model including age as a continuous predictor.

forward selection, only the parameter *D*2% of the hippocampi was selected, see figure 4.8A and table 4.8.

Memory impairment grade $\geq$ 2 was associated to *D*2% of the cerebellum (p = 0.024) and $V25Gy(RBE)$ of the brain-CTV (p = 0.049). The model including $V25Gy(RBE)$ revealed the highest *AUC* value (0.70 [0.52 – 0.88]), see figure 4.8B and table 4.8. No clinical cofactor was significantly associated to this endpoint.

No relation between dosimetric parameters and fatigue (grade $\geq$ 1 and $\geq$ 2), headache, optic nerve failure and dry eye syndrome at 12 months after PBT was observed.

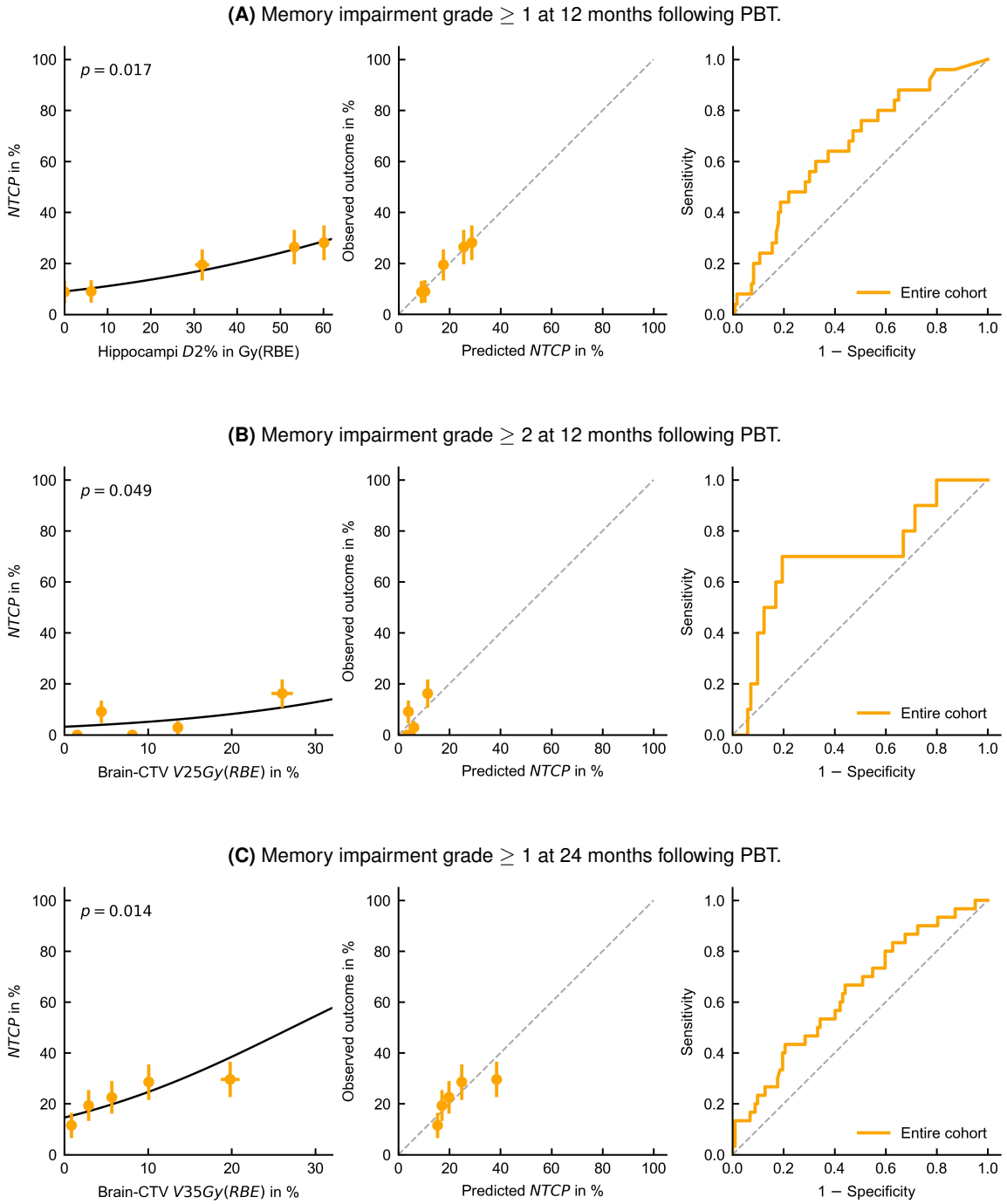


Figure 4.8: NTCP models for memory impairment grade $\geq$ 1 and grade $\geq$ 2 at 12 months following PBT **(A, B)** and grade $\geq$ 1 at 24 months following PBT **(C)** in the pooled analysis. Regression curve (left), calibration plot (centre) and ROC curve (right) are displayed. Each data point and error bar represents the mean value and standard deviation of each patient set, see section 4.1.

NTCP models for side effects at 24 months after PBT Alopecia grade ≥ 1 was associated to several dosimetric parameters of the skin ($D2\%$, $D5\%$, $D15\%$, $V10Gy(RBE)$, ..., $V50Gy(RBE)$, $p < 0.02$). The volume parameter $V30Gy(RBE)$ of the skin revealed the highest AUC value (0.83 [0.74 – 0.92], $p < 0.001$). This side effect was also associated to the clinical cofactors CTx, tumour volume (CTV), prescribed dose, and tumour location in brain or skull base ($p < 0.011$). In multivariable modelling including $V30Gy(RBE)$ and these clinical cofactors using forward selection, only the dosimetric parameter remained. Model characteristics are given in table 4.8.

Mild fatigue was associated to the near-maximum dose $D2\%$ of the brain stem (AUC = 0.62 [0.53 – 0.72], p = 0.018). The clinical cofactors CTx (p = 0.033) as well as tumour location in brain or skull base (p = 0.027) were also associated to fatigue grade ≥ 1. A bivariable model including CTx and $D2\%$ of the brain stem revealed the highest AUC value (0.68 [0.58 – 0.77]), see figure 4.9 and table 4.8.

Mild hearing impairment was significantly associated to D_{mean} (p = 0.001) and $D2\%$ (p = 0.003) of the ipsilateral cochlea in logistic regression. The model including the mean cochlear dose revealed the highest AUC value (0.79 [0.68 – 0.90]). A weak association to the clinical cofactor age at treatment was observed for hearing impairment at that time (p = 0.078). The model is presented in figure 4.7B and table 4.8.

Physician-rated memory impairment grade ≥ 1 was significantly related to dosimetric parameters of the brain-CTV ($V10Gy(RBE)$, ..., $V50Gy(RBE)$ and D_{mean}, $p < 0.035$). The model including $V35Gy(RBE)$ revealed the highest AUC value (0.64 [0.52 – 0.75], p = 0.014), see figure 4.8C and table 4.8. No significant association between mild memory impairment and any clinical cofactors was observed at that time.

No relation between dosimetric parameters and headache as well as optic nerve failure grade ≥ 1 at 24 months after PBT was observed in logistic regression.

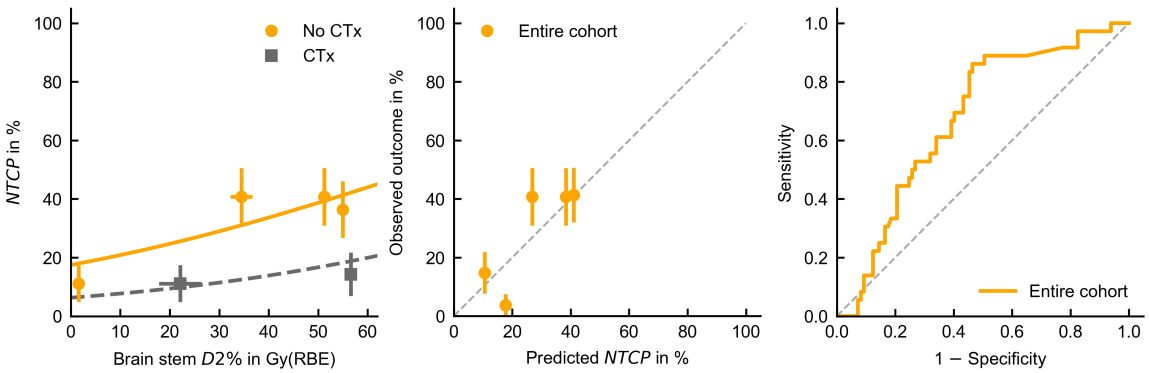


Figure 4.9: NTCP model for fatigue grade ≥ 1 at 24 months following PBT. Regression curve (left), calibration plot (centre) and ROC curve (right) are displayed. Each data point and error bar represents the mean value and standard deviation of each patient set, see section 4.1.

Generalised estimating equations for late side effects

GEEs on the pooled cohort were created to analyse all recorded patient data for all follow-up examinations between 6 and 24 months after PBT. Therefore, in addition to the dosimetric or clinical parameters, the GEEs include time as an additional parameter. The equations were created for all side effects shown in table 4.5 for grade ≥ 1 and grade ≥ 2, respectively.

Alopecia Significant associations between alopecia grade ≥ 1 and all considered dosimetric parameters of the skin were observed ($p < 0.04$), except for $V60Gy(RBE)$. The parameter $D2\%$ revealed the lowest quasi-likelihood estimation and p-value ($p < 0.001$) among all dosimetric parameters of the skin. This parameter $D2\%$ remained significant for the GEE that also includes the interaction term between time and $D2\%$ ($p < 0.001$), see table 4.9. A significant association with the endpoint was observed for the clinical cofactors prescribed dose and tumour volume (CTV) in GEEs including and excluding the interaction term between time and the respective cofactor ($p < 0.02$), see appendix B table B.12.

Several dosimetric parameters showed a significant association to alopecia grade ≥ 2 ($D2\%$, $D5\%$, $D15\%$, $D25\%$, $V10Gy(RBE)$, ..., $V50Gy(RBE)$, $p < 0.04$). The lowest quasi-likelihood estimation was observed for parameter $D2\%$ ($p = 0.012$) in both GEEs including and excluding the interaction term between time and $D2\%$, see table 4.9. A significant association with the endpoint was observed for the clinical cofactors prescribed dose and tumour volume in GEEs including and excluding the interaction term between time and the cofactor ($p < 0.03$), see appendix B table B.12.

Dry eye syndrome A significant association between mild dry eye syndrome grade ≥ 1 and D_{mean} of the ipsilateral lacrimal gland was observed in a GEE excluding the interaction term with time and this dosimetric parameter, see table 4.9. Also for dry eye syndrome grade ≥ 2 a significant association to D_{mean} of the ipsilateral lacrimal gland was observed in a GEE excluding and including the corresponding interaction term with time, see table 4.9. Moreover, dry eye syndrome grade ≥ 2 was significantly associated with the tumour location in brain or skull base in GEEs ($p < 0.001$), see appendix B table B.12.

Fatigue A significant association between mild fatigue grade ≥ 1 and two dosimetric parameters were observed in GEEs excluding the interaction term of the respective parameter and time: $D2\%$ of the brain stem ($p = 0.028$) and $D2\%$ of the cerebellum ($p = 0.034$), see table 4.9. There was no relation between fatigue grade ≥ 1 and clinical cofactors. Neither dosimetric nor clinical parameters were found to be associated with fatigue grade ≥ 2.

Headache A relation between headache grade ≥ 1 and age at treatment was observed ($p = 0.010$). Younger age was associated with milder headaches, see appendix B table B.12. How-

Table 4.9: GEE analysis for late side effects in the pooled analysis. The parameter time is defined as a natural number representing the follow-up visit starting with 1 at 6 months following PBT.

Excluding interaction term between dosimetric parameter and time				Including interaction term between dosimetric parameter and time			
Model parameter	β_i	**(95 % CI)**	**p-value**	**Model parameter**	β_i	**(95 % CI)**	**p-value**
Alopecia grade ≥ 1							
Skin $D2\%$ in Gy(RBE)$^{-1}$	0.088	(0.07 – 0.11)	< 0.001	Skin $D2\%$ in Gy(RBE)$^{-1}$	0.079	(0.04 – 0.12)	< 0.001
Time	-0.037	(-0.07 – 0.00)	0.048	Time	-0.062	(-0.17 – 0.05)	0.27
Constant	-3.33	(-4.24 – -2.42)		Interaction	0.001	(0.00 – 0.00)	0.63
				Constant	-2.99	(-4.66 – -1.33)	
Alopecia grade ≥ 2							
Skin $D2\%$ in Gy(RBE)$^{-1}$	0.10	(0.02 – 0.18)	0.012	Skin $D2\%$ in Gy(RBE)$^{-1}$	0.25	(0.05 – 0.45)	0.012
Time	-0.013	(-0.10 – 0.08)	0.79	Time	0.43	(-0.08 – 0.94)	0.10
Constant	-7.62	(-10.89 – -4.36)		Interaction	-0.009	(-0.02 – 0.00)	0.086
				Constant	-15.28	(-25.51 – -5.05)	
Dry eye syndrome grade ≥ 1							
Lacrimal gland ipsi D_{mean} in Gy(RBE)$^{-1}$	0.032	(0.00 – 0.06)	0.038	Lacrimal gland ipsi D_{mean} in Gy(RBE)$^{-1}$	0.017	(-0.01 – 0.05)	0.27
Time	-0.038	(-0.08 – 0.01)	0.11	Time	-0.054	(-0.11 – 0.00)	0.041
Constant	-2.41	(-3.28 – -1.54)		Interaction	0.001	(0.00 – 0.00)	0.31
				Constant	-2.20	(-2.97 – -1.42)	
Dry eye syndrome grade ≥ 2							
Lacrimal gland ipsi D_{mean} in Gy(RBE)$^{-1}$	0.053	(0.02 – 0.09)	0.003	Lacrimal gland ipsi D_{mean} in Gy(RBE)$^{-1}$	0.045	(0.01 – 0.08)	0.006
Time	-0.042	(-0.12 – 0.04)	0.31	Time	-0.056	(-0.16 – 0.04)	0.27
Constant	-3.93	(-5.47 – -2.38)		Interaction	0.001	(0.00 – 0.00)	0.62
				Constant	-3.74	(-5.35 – -2.13)	
Fatigue grade ≥ 1							
Brain stem $D2\%$ in Gy(RBE)$^{-1}$	0.015	(0.00 – 0.03)	0.028	Brain stem $D2\%$ in Gy(RBE)$^{-1}$	0.016	(0.00 – 0.03)	0.094
Time	-0.015	(-0.04 – 0.01)	0.24	Time	-0.012	(-0.06 – 0.03)	0.62
Constant	-1.23	(-1.86 – -0.60)		Interaction	0.000	(0.00 – 0.00)	0.88
				Constant	-1.27	(-2.05 – -0.49)	
Hearing impairment grade ≥ 1							
Cochlea ipsi $D2\%$ in Gy(RBE)$^{-1}$	0.027	(0.01 – 0.05)	0.007	Cochlea ipsi $D2\%$ in Gy(RBE)$^{-1}$	0.022	(-0.01 – 0.05)	0.16
Time	-0.006	(-0.05 – 0.04)	0.79	Time	-0.018	(-0.09 – 0.06)	0.64
Constant	-2.93	(-4.03 – -1.83)		Interaction	0.000	(0.00 – 0.00)	0.67
				Constant	-2.76	(-4.14 – -1.39)	
Memory impairment grade ≥ 1							
Brain-CTV $V25Gy(RBE)^*$	7.13	(3.88 – 10.38)	< 0.001	Brain-CTV $V25Gy(RBE)^*$	8.17	(1.85 – 14.49)	0.011
Time	0.017	(-0.01 – 0.05)	0.25	Time	0.026	(-0.02 – 0.08)	0.29
Constant	-2.43	(-3.11 – -1.76)		Interaction	-0.069	(-0.40 – 0.27)	0.69
				Constant	-2.57	(-3.53 – -1.61)	
Hippocampi $D2\%$ in Gy(RBE)$^{-1}$	0.018	(0.00 – 0.03)	0.016	Hippocampi $D2\%$ in Gy(RBE)$^{-1}$	0.019	(-0.01 – 0.04)	0.13
Time	0.017	(-0.01 – 0.05)	0.27	Time	0.019	(-0.02 – 0.06)	0.39
Constant	-2.06	(-2.82 – -1.31)		Interaction	0.000	(0.00 – 0.00)	0.92
				Constant	-2.09	(-3.02 – -1.16)	
Cerebellum $V20Gy(RBE)^*$	2.12	(0.50 – 3.73)	0.010	Cerebellum $V20Gy(RBE)^*$	1.23	(-1.60 – 4.06)	0.39
Time	0.016	(-0.01 – 0.04)	0.28	Time	0.010	(-0.02 – 0.04)	0.54
Constant	-1.71	(-2.28 – -1.15)		Interaction	0.063	(-0.14 – 0.27)	0.55
				Constant	-1.63	(-2.21 – -1.06)	
Memory impairment grade ≥ 2							
Brain-CTV $V25Gy(RBE)^*$	7.44	(2.96 – 11.92)	0.001	Brain-CTV $V25Gy(RBE)^*$	6.43	(-2.29 – 15.15)	0.15
Time	0.004	(-0.05 – 0.06)	0.88	Time	-0.007	(-0.10 – 0.09)	0.89
Constant	-3.73	(-4.89 – -2.56)		Interaction	0.066	(-0.38 – 0.51)	0.77
				Constant	-3.56	(-5.26 – -1.86)	
Hippocampi D_{mean} in Gy(RBE)$^{-1}$	0.037	(0.00 – 4.61)	0.032	Hippocampi D_{mean} in Gy(RBE)$^{-1}$	0.005	(-0.05 – 0.07)	0.86
Time	0.000	(-0.05 – 0.00)	0.99	Time	-0.027	(-0.09 – 0.03)	0.38
Constant	-2.96	(-4.07 – 27.13)		Interaction	0.002	(0.00 – 0.01)	0.18
				Constant	-2.57	(-3.63 – -1.50)	

Abbreviations: CI, confidence interval; ipsi, ipsilateral; *as fraction of the total volume.

ever, an increased dose was not associated with headache in any dosimetric parameter. Neither dosimetric nor clinical parameters were found to be associated with headache grade ≥ 2.

Hearing impairment The dosimetric parameters D_{mean} and $D2\%$ of the ipsilateral cochlea were significantly associated to hearing impairment grade ≥ 1 ($p < 0.012$) in GEE without the interaction term with time, see table 4.9. In addition, a relation between hearing impairment grade

$\geq$ 1 and age at treatment was seen in GEEs including and excluding the interaction term between time and the cofactor ($p < 0.034$), see appendix B table B.12. Older age was associated with increased hearing impairment. No association between hearing impairment grade $\geq$ 2 and dosimetric parameters of the ipsilateral cochlea was observed.

Memory impairment Several dosimetric parameters of different brain structures were significantly associated with memory impairment grade $\geq$ 1 in GEEs excluding the interaction term with time: brain-CTV ($D2\%$, D_{mean}, $V10Gy(RBE)$, ..., $V50Gy(RBE)$, $p < 0.004$), bilateral hippocampi ($D2\%$, D_{mean}, $p < 0.016$) and cerebellum ($D2\%$, D_{mean}, $V20Gy(RBE)$, $V30Gy(RBE)$, $V40Gy(RBE)$, $p < 0.05$). For each of these OARs, the model with the lowest quasi-likelihood estimate is presented in table 4.9. However, only dosimetric parameters of the brain-CTV remained significantly associated with memory impairment grade $\geq$ 1 in GEEs including the interaction term with time. Among these, the parameter $V25Gy(RBE)$ revealed the lowest quasi-likelihood estimation, see table 4.9. A significant association with memory impairment grade $\geq$ 1 was observed for the clinical cofactors prescribed dose and tumour volume (CTV) in GEEs including and excluding the interaction term between time and the cofactor ($p < 0.042$), see appendix B table B.12.

Significant associations between memory impairment grade $\geq$ 2 and D_{mean} of the bilateral hippocampi (p = 0.032) as well as certain dosimetric parameters of the brain-CTV ($D2\%$, D_{mean}, $V10Gy(RBE)$, ..., $V50Gy(RBE)$, $p < 0.014$) were observed in GEEs excluding the interaction term with time. Among these, the parameter $V25Gy(RBE)$ revealed the lowest quasi-likelihood estimation, see table 4.9.

Optic nerve failure No significant association between optic nerve failure and any parameter could be observed.

4.3.2 Discussion

This study on adult patients with intracranial tumours receiving PBT investigated the relation of late side effects within two years after PBT and dose to associated OAR. Overall, PBT was well tolerated with very low incidences of side effects grade $\geq$ 3 ($<$ 5 %). Due to these low incidence rates, NTCP models for mostly mild forms of the side effects could be created (grade $\geq$ 1). Mild alopecia was associated to doses of the skin at both 12 and 24 months after PBT. Mild hearing impairment at 24 months after PBT was associated to the mean dose of the ipsilateral cochlea. These three models could be validated on an external cohort and were also observed in the analysis of the pooled cohort. Additionally, dose-response relations were found for physician-rated memory impairment and intermediate to high doses to the remaining brain as well as high doses to the hippocampi. Moreover, mild fatigue at 24 months after PBT was associated to high doses to the brain stem as well as CTx when analysing the pooled cohort. In GEE analysis

including patient data from all available follow-up visits, the findings regarding alopecia as well as memory impairment could be confirmed. Additionally, dry eye syndrome grade $\geq$ 2 was associated to the mean dose of the ipsilateral lacrimal gland.

Incidence rates of late side effects

Although many patients have already been treated with PBT, overall published data on incidence rates of physician-rated side effects with a follow-up period of more than one year are rare. Shih et al. (2015) investigated various late side effects in 20 LGG patients treated with 54 Gy(RBE) and a DS PBT technique. During follow-up (median: 5.1 years), the most common physician-rated side effects (CTCAE) after more than three months following treatment were persistent headaches, fatigue, alopecia and new neurologic deficits. These findings are generally comparable to those in this thesis. However, mild headache was less frequent in the present study (10 – 15 %) compared to the study by Shih et al. (2015) (60 %). Similarly, less fatigue grade 3 was reported in the present thesis (2 %) compared to the discussed study (15 %). However, the follow-up in this thesis was somewhat shorter, so that more severe toxicities may develop later on.

Weber et al. (2012) assessed the long-term outcome of 39 patients with intracranial meningiomas treated with normo-fractionated PBS PBT and a median dose of 56 Gy(RBE). The mean follow-up time was 5.2 years. Late side effects according to CTCAE were observed in 41 % of the patients. Pituitary dysfunction was most common (10 %), followed by transient brain oedema (8 %), optic nerve neuropathy (8 %), brain necrosis (5 %), and retinitis (5 %). 31.3 % of the patients presented with late side effects grade $\geq$ 3, although these patients were not pre-irradiated. In comparison, in the present thesis, the incidence rate for side effects grade $\geq$ 3 was well below 30 %, see table 4.6. Optic nerve failure occurred with a comparable frequency. In the present study, it ranged between 2 and 8 % (grade $\geq$ 2). However, it should also be noted that the present study only had a follow-up period of 2 compared to 5 years in the study by Weber et al. (2012). Moreover, several of the before-mentioned side effects observed by Weber et al. (2012) were not investigated in all patients in this thesis, such as pituitary dysfunction or retinitis.

Noël et al. (2005) assessed the outcome of 51 patients with intracranial benign meningiomas located in the skull base that were treated with a combination therapy of normo-fractionated XRT and PBT and a median dose of 60.6 Gy(RBE). During follow-up (mean: 2.1 years), one patient (2 %) reported grade 3 hearing loss and another a complete loss of pituitary function (grade 3 according to LENT-SOMA). However, they did not report mild or moderately severe clinical side effects. The follow-up time was comparable between this study and the presented study, as was the incidence rate of grade 3 hearing loss (3 %).

A study on 77 skull-base chondrosarcoma patients treated with PBS PBT and doses between 64.0 – 76.0 Gy(RBE) observed high-grade (grade $\geq$ 3) radiation-induced side effects in 8 % of the patients including hearing loss, radiation necrosis and optic nerve neuropathy during a mean follow-up of 5.8 years (Weber et al., 2016). They reported incidence rates of any side effects

according to CTCAE but did not further specify the type of toxicity (grade 1: 21 %, grade 2: 30 %, grade 3: 4 %, grade 4: 4 %).

In general, the incidence rates of late effects in this thesis did not differ markedly from those reported in the other studies mentioned above.

Dose-response relationships for late side effects

In this thesis, dose-response relationships for several late side effects following PBT were observed. In the following, the findings are compared with those of other dose-response studies.

Alopecia Mild alopecia was related to high dose parameters of the skin ($D2\%$, $V30\,Gy(RBE)$ and $V45\,Gy(RBE)$) at 12 and 24 months after PBT. High *AUC* values were observed in exploration and validation. Palma et al. (2020) investigated 116 adult brain tumour patients treated with PBS PBT with a median dose of 54 Gy(RBE). Using the CTCAE grading system, late ($>$ 90 days after PBT) and permanent ($>$ 12 months after PBT) radiation-induced alopecia grade $\geq$ 2 was observed in 35 % and 19 % of the patients, respectively. These incidence rates are considerably higher than those observed in this thesis as only 2 % of the patients in the exploration cohort presented with alopecia grade $\geq$ 2 at 12 and 24 months after PBT. Palma et al. (2020) developed NTCP models based on dose-surface histogram parameters. An LKB model showed increasing tolerance values from 24 Gy(RBE) for late alopecia to 44 Gy(RBE) for permanent alopecia. The effect of increasing tolerance doses over time was also observed in this thesis, as described below in more detail. The most predictive dosimetric parameter for late alopecia grade $\geq$ 2 characterised the skin surface receiving 25 Gy(RBE). Surgery was a risk factor for late alopecia grade $\geq$ 2 as it was in this thesis for late alopecia grade $\geq$ 1. Palma et al. (2020) found $D2\%$ of the scalp as the only predictor for permanent alopecia. This parameter was also found to be predictive for alopecia grade $\geq$ 1 in the presented thesis. Moreover, this factor seems to be predictive for both PBT techniques, PBS and DS.

Lawenda et al. (2004) investigated permanent alopecia ($>$ 12 months) after cranial XRT with prescribed doses between 30.6 and 63 Gy in 26 patients. Permanent alopecia was correlated to dose to the hair follicles. In logistic regression, the tolerance dose TD_{50} for alopecia grade $\geq$ 3 was found to be 43 Gy. Severity grading was not based on standardised scoring systems, such as CTCAE, but on a 4-point scale with grade 1 *none*, grade 2 *minimal*, grade 3 *moderate* and grade 4 *severe*. However, the CTCAE grading for alopecia ranges from 0 to 2 as the most severe grade, see appendix B table B.13. Thus, alopecia grade 3 in the study by Lawenda et al. (2004) may approximately correspond to alopecia grade 1 according to CTCAE. In the logistic models developed in this thesis for alopecia grade $\geq$ 1 at 12 months after PBT including $D2\%$ of the skin, the tolerance dose TD_{50} was estimated to 41.2 Gy(RBE) in the exploration cohort (see 4.4A and table 4.7) and 42.7 Gy(RBE) in the pooled cohort (see table 4.8). Although different RT techniques were applied, these findings are in good agreement with those by Lawenda et al.

(2004). The authors did not find an association between alopecia and age at treatment or gender. This is in accordance with the findings of this thesis.

Hair loss is often considered a minor unavoidable side effect of cranial RT (Shih et al., 2009). However, it may have a great impact on patients' QoL, even years after the initial event. This is reflected in reduced self-confidence and low self-esteem. These concerns could be addressed by creating RT treatment plans that could reduce the probability of developing late alopecia.

Hearing impairment In this thesis, mild hearing impairment was found to be associated to the mean dose to the ipsilateral cochlea as well as age at PBT. Other studies reported similar causative factors in patients treated with XRT. Bhandare et al. (2010) published a comprehensive review of literature about hearing loss following XRT or stereotactic RT for patients with head-and-neck cancer or vestibular schwannoma. Several treatment- and patient-related factors to hearing loss were revealed. Among treatment-related factors, the mean dose to the cochlea during fractionated XRT was found to be important as in this thesis.

Hearing impairment 2 to 17 years after treatment was assessed by van der Putten et al. (2006) in 21 patients with unilateral tumours of the parotid gland who underwent parotidectomy and postoperative XRT. The contralateral ear was used as a reference in pure-tone and speech audiometry. A hearing loss of more than 15 dB in 3 frequencies was found in patients with mean doses to the cochlea and Eustachian tube of more than 50 Gy. They also fitted an NTCP model with a 10 % probability for asymmetric hearing loss of 15 dB for a mean dose to the inner ear of 42 Gy. However, they did not report the model coefficients. In 31 head-and-neck patients treated with normo-fractionated 3D-CRT, Pan et al. (2005) investigated the differences in threshold level between the ipsilateral and contralateral ears on pure-tone audiograms up to 3 years after XRT. Hearing loss was related to the mean cochlear dose: a loss of $\geq$ 10 dB was observed in patients with a mean cochlear dose of more than 45 Gy. The authors developed a linear model and presented a nomogram that included the test frequency, age at treatment, baseline hearing, and the dose differences between the ipsi- and contralateral inner ear.

The model developed in this thesis predicts a 10 % probability for physician-rated hearing impairment grade $\geq$ 1 for a mean cochlear dose of 21.8 Gy(RBE), see table 4.7. Grade 1 hearing impairment is defined in the CTCAE for adults that were enrolled on a monitoring program as a threshold shift of 15 – 25 dB averaged at 2 contiguous test frequencies in at least one ear. As the patients enrolled in this thesis did not undergo such assessments using audiograms, CTCAE grade 1 refers to a *subjective change in hearing*. The different methods of assessment may explain the differences in model predictions and tolerance doses.

More recently, De Marzi et al. (2015) investigated 114 adult patients with chordoma and chondrosarcoma of the skull base treated with postoperative 3D-CRT and PBT with a median dose of 70 Gy(RBE) and a minimum follow-up of 26 months. Hearing impairment was defined as hearing loss $>$ 15 dB at two contiguous test frequencies, tinnitus or otitis media and recorded for each ear separately. Patients with hearing impairment had a cochlear mean dose of 54.6 Gy(RBE)

while this value was only 36.8 Gy(RBE) in patients without any effect. They developed a logistic model including cochlear mean dose as model predictor with TD_{50} = 56.0 Gy(RBE). The NTCP model predicting hearing impairment grade $\geq$ 1 at 24 months after PBT developed on the exploration cohort in this thesis (see table 4.7) had a TD_{50} = 79.4 Gy(RBE). The logistic model for the same endpoint on the pooled cohort (see table 4.8) had a TD_{50} = 69.4 Gy(RBE). These higher values of TD_{50} compared to the study by De Marzi et al. (2015) may be explained by the different assessment of toxicity. They analysed a combined endpoint (hearing loss $>$ 15 dB at two contiguous test frequencies, tinnitus or otitis media), while in this thesis, hearing impairment was assessed by physicians according to the symptom *hearing impairment* as defined in the CTCAE scoring system. Moreover, the follow-up period was shorter in this thesis compared to the study by De Marzi et al. (2015). The authors may have detected hearing defects that may occur more than 2 years after treatment but at lower cochlear doses.

Hearing loss was found to be associated to dose to the vestibulocochlear nerve in stereotactic RT (Bhandare et al., 2010). However, this relation could not be analysed in this thesis, as the eighth cranial nerve was not delineated. A possible synergistic effect of RT and CTx with cisplatin to hearing loss was described for patients with head-and-neck cancer (Chen et al., 2006; Bhandare et al., 2010). However, in brain tumour patients, this type of CTx drug is of minor importance compared to other agents such as TMZ or PCV (Roth and Weller, 2015). Moreover, the incidence rate of hearing loss after XRT increases with age ($>$ 50 years) (Ho et al., 1999; Honoré et al., 2002; Pan et al., 2005; Bhandare et al., 2007; Bhandare et al., 2010). This was also observed in the present study on patients treated with PBT. One study observed higher incidence rates of hearing loss after RT in men compared to women (Kwong et al., 1996). However, this effect could not be confirmed in this thesis, nor in other studies (Bhandare et al., 2007).

As mentioned before, in this thesis, hearing loss was assessed by physicians, who had to base their assessment on the patients' explanations. As a result, the assessment strongly depended on the type of questioning and interpretation, which gives the grading a more subjective character. No audiograms were recorded that would allow an accurate assessment of the hearing condition (Ho et al., 1999; Pan et al., 2005; Chen et al., 2006). The CTCAE scoring system provides a grading based on audiogram data (referred to as *Adults enrolled on a Monitoring Program*). Bhandare et al. (2010) have summarised comprehensive recommendations for scoring ototoxicity including pre- and post-treatment evaluations of the same ear using audiograms, whereby the evaluation should be performed first 6 months after RT and thereafter at least every 6 months. Several important test frequencies are also presented there. For future studies, a standardised measurement using pre-and post-treatment audiograms is recommended.

Memory impairment Memory impairment was related to intermediate and high doses to the remaining brain as well as to high doses to the hippocampi in the pooled cohort. Fairly low *AUC* values ($<$ 0.7) were observed and these models need to be independently validated. Gondi et al. (2012) investigated patients with benign or low-grade brain tumours and found a signifi-

cant relationship between delayed recall on the Wechsler Memory Scale-III Word List (WMS-WL) at 18 months after XRT and the dosimetric parameter $D40\%$ of the bilateral hippocampi with TD_{50} = 14.88 Gy. In this thesis, the model for physician-rated memory impairment grade $\geq$ 1 at 12 months after PBT including $D2\%$ of the hippocampi as model predictor (see table 4.8) had a much higher TD_{50} = 100.9 Gy(RBE). Physician-rated memory impairment according to CTCAE was assessed in this thesis. The specifications of the individual severity grades are very imprecisely formulated, e.g. CTCAE grade 1 is defined as *mild memory impairment*, grade 2 *moderate memory impairment with limiting instrumental activities of daily living* and grade 3 *severe memory impairment with limiting self-care activities of daily living*. This clearly shows that the assessment depends very much on the interpretation of the patient's answers by the physicians. Due to the rather unspecific nature of this endpoint, it is challenging to find dose-response relationships for physician-rated memory impairment. Hence, a more detailed investigation of neurocognitive deficits using specific tests is required. In chapter 5, a comprehensive analysis and discussion of neurocognitive function after cranial PBT is presented.

Fatigue At 24 months after PBT, mild fatigue was associated to the near-maximum dose of the brain stem as well as CTx when analysing the pooled cohort. This model revealed an *AUC* value of 0.68 and requires independent validation. As discussed in section 4.2, in the PARSPORT Phase III trial including patients with pharyngeal squamous-cell carcinoma, patients presenting with acute fatigue grade $\geq$ 2 had higher maximum and mean doses in the posterior fossa, brain stem and cerebellum compared to patients with acute fatigue grade $\leq$ 1 (Gulliford et al., 2012a). However, data on late fatigue were not acquired in this trial, as it was not expected to be a late side effect of RT (Nutting et al., 2011). Nevertheless, radiation-induced fatigue is a common chronic side effect described by 30 % of the patients during follow-up (Jereczek-Fossa et al., 2002). Similarly, in this thesis, fatigue was one of the most common late side effects, as in the study by Shih et al. (2015). Self-reported fatigue was even found to be a predictor of survival of patients with recurrent HGG (Peters et al., 2014). Due to the fact that chronic fatigue after cranial RT is often underestimated by physicians, there are almost no studies on dose-effect relationships for this side effect. Therefore, the prevalence and predictive factors of chronic fatigue are not really understood (Jereczek-Fossa et al., 2002). Fatigue is potentially a synergistic effect of many influencing factors. These could include biochemical factors, psychological disturbances such as depression or anxiety, further therapies such as CTx, hormone therapy or surgery or comorbidities such as pain or electrolytic disturbances (Jereczek-Fossa et al., 2002). In a study on breast cancer patients, women treated with CTx presented with higher fatigue than those who were not treated with CTx (Bower et al., 2011). In this thesis, however, patients treated with CTx were found to present with lower fatigue than others. This was also observed in a study on lung cancer patients, in which a lower frequency of fatigue was observed in patients with previous surgery or CTx (Hickok et al., 1996). In a study of acute fatigue in head-and-neck cancer patients, fatigue did not appear to be caused by CTx (Ferris et al., 2018). Because of these controversies,

the relationships between fatigue and potential predictive factors should be investigated in larger cohorts including QoL questionnaires.

Dry eye syndrome Dry eye syndrome grade $\geq$ 2 was associated to the mean dose of the ipsilateral lacrimal gland in GEE analysis of the pooled cohort, which requires independent validation. The finding is in accordance with other studies. Bhandare et al. (2012) retrospectively assessed severe dry eye syndrome leading to vision compromise in 78 patients with extracranial head-and-neck tumours treated with XRT between 1965 and 2000. The endpoint was assessed using one or more specific tests that measure the amount, instability and changes in the tear film. The incidence of dry eye syndrome increased with increasing dose to the lacrimal gland. The authors developed an NTCP model with TD_{50} = 48.0 Gy. No association was seen to CTx, age at treatment, and gender. This is in accordance with the findings of this thesis.

Batth et al. (2013) retrospectively investigated dose-response relations between dose to the lacrimal gland and ocular toxicity in patients with sinonasal tumours treated with a median dose of 66 Gy using IMRT. Ocular toxicity was defined as conjunctivitis, corneal ulceration and keratitis according to the RTOG scoring system. During a median follow-up of 28.5 months, the incidence rates of late ocular toxicity of grades 1, 2, and 3 were 10 %, 8 % and 13 %, respectively. In the presented thesis, the incidence rates in the exploration cohort were lower with no grade 3 and 6 % and 4 % for grade 1 and grade 2, respectively, see table 4.6. No events of dry eye syndromes were observed in the validation cohort. The differences in the incidence rates may be explained by the different tumour histologies and locations as well as the higher prescribed doses in the study by Batth et al. (2013). The authors developed a logistic NTCP model for late ocular toxicity grade $\geq$ 1 including the maximum dose to the lacrimal gland as predictor but did not report all model coefficients. With an increase of the maximum dose by 1 Gy, the odds of developing late ocular toxicity grade $\geq$ 1 increased by 7 %.

As dry eye syndrome has a negative impact on patients' QoL (Grubbs et al., 2014), a comprehensive assessment of dry eye syndrome following PBT could be included in future studies. Additionally, specific QLQs could be used as the brain-tumour specific module of the EORTC (QLQ-BN20) does not address this side effect.

Optic nerve failure In this thesis, no association between optic nerve failure and dose to the optic nerve or chiasm was observed. This may be due to the low incidence rates in the observed follow-up period. The latency period between RT and visual loss due to radiation-induced optic neuropathy may range from three months to 8 years (Danesh-Meyer, 2010). Generally, it is often reported that doses of more than 50 Gy or 60 Gy are necessary to induce radiation-induced optic neuropathy (Danesh-Meyer, 2010; Mayo et al., 2010b).

Dose-response over time

Mild alopecia was associated to high dose parameters of the skin, both at 12 months and at 24 months after PBT. All analyses showed that especially the near-maximum parameter $D2\%$ was significantly related to this side effect at both time points. For this parameter, figure 4.10A shows the logistic models for alopecia grade ≥ 1 at both time points for the pooled cohort. The curve is slightly steeper at 12 months compared to 24 months after PBT, which is also apparent in the different model coefficients (0.12 vs 0.08 Gy(RBE)$^{-1}$, respectively) and the increasing tolerance doses TD_{50} (41.2 vs 48.7 Gy(RBE), respectively). In GEE analysis including all follow-up examinations up to 24 months after treatment, the parameter time had a negative coefficient ($p = 0.048$), which indicates a slight decrease in the incidence of the side effect over time. A similar effect for alopecia grade ≥ 2 was observed by Palma et al. (2020) when comparing late and permanent alopecia.

Mild hearing impairment was associated to the mean dose of the ipsilateral cochlea, both at 12 months and at 24 months after PBT in the analysis of the pooled cohort. Figure 4.10B shows the logistic models for hearing impairment grade ≥ 1 at both time points. In contrast to mild alopecia, the curve is slightly shallower at 12 months than at 24 months after treatment, which is also reflected in the model coefficients (0.035 vs 0.050 Gy(RBE)$^{-1}$, respectively) and the decreasing tolerance doses TD_{50} (89.5 vs 69.8 Gy(RBE), respectively). This indicates that the probability of developing mild hearing deficits at the same mean cochlear dose increases over time. This finding is consistent with other studies analysing patients treated for nasopharyngeal carcinoma with 3D-CRT, IMRT and electrons (Ho et al., 1999; Chen et al., 2006). In both studies, pre- and

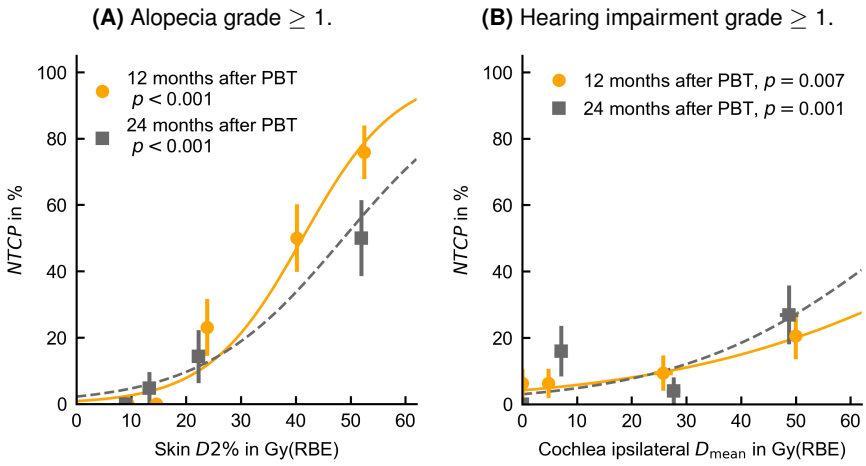


Figure 4.10: NTCP models for alopecia grade ≥ 1 **(A)** and hearing impairment grade ≥ 1 **(B)** at 12 and at 24 months following PBT on the pooled cohort for the same model predictors, respectively. Each data point and error bar represents the mean value and standard deviation of each patient set, see section 4.1.

post-RT pure-tone audiograms were performed to assess the extent of hearing loss and they concluded that permanent hearing loss increases with time after treatment.

For the model-based selection of patients for PBT, comprehensive NTCP profiles could be used that predict complication risks at multiple time points after treatment ideally including the same model predictors (Van den Bosch et al., 2019).

Side effects not investigated in this thesis

Endocrine dysfunction due to irradiation of the hypothalamic–pituitary axis is a side effect that occurs after cranial RT in children and adults (Darzy and Shalet, 2009). In adults, endocrine dysfunction may develop over months to years and is often not diagnosed. Most hormone deficiencies can usually be corrected by pharmacotherapy (Shih et al., 2009). In a study by Shih et al. (2015), 20 LGG patients treated with PBT at a dose of 54 Gy(RBE) underwent endocrine assessment during follow-up (median 5.1 years). In 6 patients a new endocrine dysfunction was detected. De Marzi et al. (2015) investigated patients with chordoma or chondrosarcoma treated with a combination of 3D-CRT and PBT. Of 103 adults, 45 patients developed endocrine dysfunction. They fitted an NTCP model for endocrine dysfunction with a TD_{50} value of 60.5 Gy(RBE) for the pituitary gland. To establish accurate dose-response relationships between endocrine dysfunction after PBT and pituitary dose, endocrine monitoring of all study patients would be necessary and is recommended for future studies.

In this thesis, optic nerve failure has been assessed. However, other cranial nerves may also be affected by irradiation, which may lead to cranial neuropathies. However, these radiation-induced cranial neuropathies are rare and their occurrence may have a long latent period of several years (Kong et al., 2011). Studies on patients with nasopharyngeal carcinoma revealed several causative factors such as muscle fibrosis of the neck, total dose to the nasopharynx, CTx, and cranial nerve failure at diagnosis (Lin et al., 2002; Kong et al., 2011). Although an accurate diagnosis requires comprehensive physical examinations for each cranial nerve, future studies on long-term effects of PBT may also record cranial neuropathies.

In this thesis, physician-rated side effects have been considered. These reflect the clinical symptoms and therefore the relevance for the patients. However, assessments by third persons are always affected by interobserver variability, as the physicians need to interpret the patients' statements. In addition to physician-rated side-effects, a more objective evaluation of some RT effects can be obtained using MR images during follow-up. In glioma patients treated with XRT or PBT, white matter changes in terms of reduced diffusivity were observed in areas receiving 10 – 20 Gy(RBE) (Raschke et al., 2019) or even in areas receiving less than 10 Gy (Connor et al., 2016). These changes were dose-dependent and deteriorated with time after treatment. These image changes do not necessarily reflect the clinical condition of the patient. Nevertheless, focal image changes observed on MRI scans may develop into radiation-induced brain necrosis, a severe side effect. The development of NTCP models for radiation-induced necrosis is demand-

ing as a distinct differentiation of necrotic tissue from tumour progress is challenging. Various MRI techniques such as diffusion tensor imaging, perfusion imaging and spectroscopy are commonly used for the diagnosis of necrotic brain tissue (Viselner et al., 2019). Moreover, patients with brain necrosis may be asymptomatic or present with similar clinical symptoms as tumour progress (Verma et al., 2013). Therefore, the definition of the modelling endpoint is complex. Furthermore, necrosis presents as single or multifocal lesions involving one or more lobes in one or both hemispheres (Verma et al., 2013). Due to this very local occurrence, it is difficult to define a dedicated OAR for modelling. Radiation-induced enhancing foci are typically observed in the periventricular region (Shah et al., 2012; Eulitz et al., 2019a). However, this area has not yet been routinely contoured in clinical practice, so there are no validated models based on this structure available.

Although radiation-induced necrosis is a rare side effect, reports of brain stem necrosis after PBT indicate possible biological differences between XRT and PBT (Haas-Kogan et al., 2018). For proton beams, both LET and RBE are variable and increase in the distal edge of the spread-out Bragg-Peak (Paganetti, 2014; Lühr et al., 2018). For treatment plans with several beam directions, LET/RBE may superimpose and further increase in the overlap area leading to enhanced biological effects. For single fields, the effect at field edges may also be substantial, since the LET/RBE of several fields cannot be averaged.

These effects could be even more pronounced when moving from DS techniques to intensity-modulated PBT, as these may further alter LET and RBE (Haas-Kogan et al., 2018). There is a growing interest in understanding the development of contrast-enhancing lesions that may evolve into necrotic brain regions by exploring the capabilities of treatment plan evaluation based on LET and RBE. Models predicting the local occurrence probability of contrast-enhancing lesions is part of ongoing research (Eulitz et al., 2019a; Eulitz et al., 2019b).

Scoring of side effects of paired organs

In paired organs, such as the lacrimal gland, optic nerve or cochlea, one organ is usually exposed to higher doses compared to the other. As a result, the associated side effects may occur only on one side, e.g. dry eye syndrome, optic nerve failure or hearing impairment. However, the affected side of the respective toxicity was not recorded in all cohorts included in this thesis. Therefore, it was assumed that the severity grade of the side effect referred to the side that was exposed to a higher dose. Thus, only the respective ipsilateral organ was examined. For a more detailed analysis, it would be advisable to precisely record the side of the toxicity in future studies.

Limitations of the studies on early and late side effects

The cohorts included in this study were rather heterogeneous with regard to different tumour histologies. The exploration cohort included patients with glioblastoma who have a comparably short life expectancy. However, a separate analysis for different tumour histologies was not feasible as

the number of patients and events in each subgroup would have been even lower. An increasing number of patients is being treated within ongoing clinical trials, e.g. (NCT02824731, 2016). Once these studies will have been completed and the primary endpoint is evaluated, these patient data can be used to validate the NTCP models presented in this thesis or to develop further NTCP models.

All dosimetric parameters were based on the planned dose distribution. The differences between the planned dose and the delivered dose were not considered in this study. The planning CT is only a single snapshot that is not representative of the anatomical situation during every treatment fraction. For OARs with small volumes, such as cochlea or lacrimal gland, slight positioning errors may lead to deviations of the delivered dose from the planned dose. To evaluate the administered dose, regular CT scans must be acquired during the course of treatment to calculate the actual dose distribution. In brain tumour patients, however, this is usually not done because they can be fixed very well using tight masks. Moreover, there is practically no organ movement that would justify such frequent CT scans. It can, therefore, be assumed that the results of this study would not differ substantially if the estimated delivered dose would have been assessed.

The dosimetric parameters extracted from DVHs may not consider the spatial information of the three-dimensional dose-distribution. Thus, different parts of an OAR are treated as having equal functions but this does not necessarily have to be correct (Marks et al., 2010; Gulliford, 2015). However, this may be of little relevance for small organs such as the cochlea or lacrimal gland, but it may be more important for large organs such as the remaining brain that consists of various substructures. Therefore, in this thesis, some subregions of the brain were analysed, e.g. hippocampi or cerebellum. Due to the specific dose prescriptions for different tumour histologies, only data for certain dose ranges have been available (30 – 74 Gy(RBE)). Thus, only a part of the NTCP curve could be estimated based on these data, leading to an increased uncertainty beyond this dose range.

Another important uncertainty arises from variations in the delineation of OARs. Associated dosimetric parameters are affected, which in turn would modify the NTCP models. Although standardised delineation guidelines were used in this study, physician-individual variations may still occur, as contouring can be largely inconsistent between clinicians (Foppiano et al., 2003; Rosewall et al., 2011; Nelms et al., 2012a). However, this variability, which inevitably also occurs in clinical practice, is therefore already incorporated into the models. To reduce contouring variability, consistent use of delineation guidelines and atlases is essential (Wright et al., 2019). In clinical practice, often only essential OARs are delineated that are expected to receive considerable doses and for which published dose constraints exist. For modelling purposes, however, other structures that are not routinely contoured are relevant to generate dose-response relationships that may lead to new OARs and associated dose constraints in clinical practice. The retrospective contouring of large cohorts is a challenge. To reduce the delineation time, the implementation of auto-contouring tools could be considered (Vinod et al., 2016). These tools are

currently being further improved, e.g. by deep-learning-based segmentation approaches (van der Heyden et al., 2019; van Dijk et al., 2020).

The patients included in this thesis have been enrolled in different clinical trials. The study protocols of these trials included a physician-rated assessment of several side effects according to CTCAE. NTCP modelling was mainly limited due to the low incidence rates and severity grades of several physician-rated side effects. However, in order to further reduce the observed side effects, NTCP models and dose limits are still required. Therefore, future studies should enrol a larger number of patients, which can only be achieved in a multicentre approach.

A general issue in daily practice and clinical trials is underreporting of toxicities (Trotti and Bentzen, 2004; Bentzen et al., 2010). Furthermore, the toxicity assessment by different physicians during follow-up is subject to interobserver variability even if the same scoring system is used, see section 4.4. Especially in case of rather unspecific side effects, the grading depends on the physician's experience (e.g. communication skills and interpretation of patient's answers). It is therefore essential to hold regular training courses, achieving an equal side effect assessment in at least one institute.

A major limitation of modelling late side effects in this study were different follow-up schemes applied in the participating institutions. Thus, only patient data from two different time points (12 and 24 months after PBT) could be used for modelling and validation. In this thesis, side effects up to 2 years after therapy could be investigated. However, some side effects manifest later on, e.g. optic neuropathy (Danesh-Meyer, 2010). The longer the follow-up, the more patient data will be missing because some patients have a short life expectancy, develop a recurrence or may just not attend follow-up visits. Instead of a complete case analysis, multiple imputation methods could be used to impute both predictor variables and outcome values (Sterne et al., 2009). However, these techniques may also introduce bias, especially when imputing outcome data and thus, cannot replace high-quality data collection.

In the course of this thesis, many clinical variables were investigated that may affect the occurrence of several side effects, for example, age at treatment for hearing impairment. However, other cofactors that have not been recorded may also cause clinical effects. Especially in the case of rather unspecific side effects such as fatigue, further factors may influence the occurrence, e.g. biochemical factors, psychological disturbances or comorbidities (Jereczek-Fossa et al., 2002). If studies investigate such rather unspecific endpoints, more detailed data on the living conditions of the patients should be collected.

4.4 Interobserver variability of alopecia and erythema assessment

The quality of an NTCP model is essentially determined by the accuracy of the input parameters. For NTCP models, dosimetric parameters as well as toxicity scores serve as input values, which both may be subject to inaccuracies. The dosimetric parameters may not reflect the applied dose derived from planning CT scans. Due to varying patient positioning or changes in the anatomy,

the delivered dose may deviate *in vivo* from the planned dose. However, for brain tumour patients treated with PBT, these effects may be mostly negligible compared to other tumour entities. Patients are immobilised with rigid facial masks and anatomic changes (e.g. weight loss or gain or tumour shrinkage) are rarely present in the cranial region.

A further uncertainty may be the assessment of side effect severity. Some side effects following cranial RT are difficult to quantify, e.g. headache, concentration or fatigue. The clinician's severity estimation depends greatly on the patients' symptom description and thus on the clinician's investigative questions. Furthermore, follow-up visits are just snapshots in time, while side effect severity may vary. Additionally, the same patient may be assessed by different clinicians during the course of follow-up visits. Thus, efforts have been taken to standardise the assessment. A variety of grading systems has been developed in the last decades, see section 2.3.1. These grading systems such as the CTCAE are widely used in both clinical routine and trials. Although they offer guidelines for assessing a variety of side effects, grading scores may still vary between both individual observers and institutions. In the study on modelling early side effects of PBT, a shift in the calibration of the alopecia model was observed, most likely caused by different toxicity assessments, see section 4.2.2. As consistent toxicity assessment is crucial for modelling of side effects, this study aims to investigate the interobserver variability of erythema and alopecia using cranial photographs of brain tumour patients treated with PBT.

4.4.1 Patient cohort and experimental design

Patient data

A patient cohort of 15 patients with brain tumours treated with PBT at UPTD, who were enrolled in the clinical trial (NCT02824731, 2016), were included in this study. All patients gave informed consent. Two anonymised photographs of the external cranial irradiation area were taken for each patient in the first week and later during treatment. Every appointment was regarded as a single study case. For one patient, only one photograph could be taken in the first week of treatment, resulting in a total of 29 cases that could be analysed. All photos for all cases were merged into one single document. An example case is given in appendix B figure B.1. Twenty-three radiation oncologists from five different institutions were asked to assess all cases with regard to erythema and alopecia according to CTCAE version 4.0. The definitions of each grade for both side effects are given in appendix B table B.13. The number of physicians varied between the institutions: 9 physicians from UPTD, 3 physicians from WPE, 4 physicians from MGH, 3 physicians from the Department of Radiation Oncology of the University Hospital Tuebingen (TUEB) and 3 physicians from the Department of Radiation Oncology and Radiotherapy of the Charité Berlin (BER).

Evaluation

To evaluate the variability of the toxicity assessment between the different scorers and different institutes, percent agreement as well as other metrics were analysed (described below). In some cases, where physicians recorded two grades at a time (as may occur in clinical routine), the maximum value was used. As there is no single *true* severity grade, the modal value was calculated for each case across all clinicians. This overall modal value was considered as *true* severity grade. Additionally, the modal value of all ratings within each institute was evaluated. The incidence rates of each side effect were compared between all institutions based on these modal values. Moreover, the percent agreement with the *true* severity grade (modal value) of all physicians was assessed for each case and side effect. For each institute, the number of cases deviating from the *true* severity grade was obtained.

To measure the reliability between observers for toxicity grading as a categorical metric, Cohen's kappa κ_C was used. In contrast to percent agreement, κ_C considers that agreements can also occur randomly (Cohen, 1960). As this metric only assesses the agreement between two scorers, pairwise κ_C were calculated for all combinations of scorers and the arithmetic mean $\overline{\kappa_C}$ was evaluated as suggested by Light (1971). Additionally, the agreement of the scorers was assessed using the intraclass correlation coefficient *ICC* based on a two-way random model, single measures and absolute agreement (also called *ICC*$(2,1)$) as both patients and scorers were randomly chosen from a larger pool of potential subjects (Shrout and Fleiss, 1979). Moreover, the coefficient Krippendorff's alpha α_K was used (Krippendorff, 2004). This metric is applicable to any number of scorers and is able to handle missing data. All these metrics can reach a maximum value of 1, indicating an almost perfect agreement. Values of 0 (or less) represent an agreement that would be expected by chance. More detailed interpretation of the values has been described in different publications. While Landis and Koch (1977) suggested values in the range of 0.81 – 1.00 as *almost perfect*, 0.61 – 0.80 as *substantial*, 0.41 – 0.60 as *moderate* and 0.21 – 0.40 as *fair*, Koo and Li (2016) indicated $>$ 0.90 as *excellent*, 0.75 – 0.90 as *good*, 0.50 – 0.75 as *moderate* and $<$ 0.50 as *poor*.

In clinical practice, the severity grade is usually recorded by only one physician rather than the modal value of several physicians of an institute. Hence, the interobserver variability was additionally investigated by randomly selecting only one physician from each of the institutes. The mean value for 1000 samples was examined. The analysis was conducted in R (R Core Team, 2017) with the package irr (Gamer et al., 2010).

4.4.2 Results

Interobserver variability between all clinicians

The distribution of the scored severity grades for alopecia and erythema in terms of modal values for each of the 29 cases over all physicians is given in the first row of table 4.10. For alopecia,

Table 4.10: Distribution of severity grades for the scoring of alopecia and erythema in terms of modal values for each case over all physicians (first row) and each institute, given in per cent.

	Alopecia			Erythema			
Institute	**Grade 0**	**Grade 1**	**Grade 2**	**Grade 0**	**Grade 1**	**Grade 2**	**Grade 3**
All	27.6	37.9	34.5	27.6	72.4	0.0	0.0
WPE	34.5	55.2	10.3	27.6	62.1	10.3	0.0
MGH	24.1	34.5	41.4	20.7	69.0	10.3	0.0
UPTD	44.8	27.6	27.6	34.5	62.1	3.4	0.0
BER	24.1	48.3	27.6	24.1	72.4	3.4	0.0
TUEB	31.0	37.9	31.0	34.5	58.6	6.9	0.0

Abbreviations: WPE, West German Proton Therapy Centre Essen; MGH, Massachusetts General Hospital in Boston; UPTD, University Proton Therapy Dresden; BER, Department of Radiation Oncology and Radiotherapy of the Charité Berlin; TUEB, Department of Radiation Oncology of the University Hospital Tuebingen.

37.9 % and 34.5 % of the cases were rated with grade 1 and grade 2, respectively. No patient was rated with erythema grade 2 or grade 3 during treatment based on the modal values of all physicians. Mild erythema grade 1 was most common (72.4 % of the cases).

Figure 4.11 shows the distribution of the gradings of all physicians and for all cases sorted by severity grade (overall modal value) and percent agreement. The percent agreement for all cases was above 43.5 % with an average percent agreement of 69.3 % and 73.2 % for alopecia and erythema, respectively. The distribution of the percent agreement values is displayed in appendix B figure B.2. For some cases, the assessment differed between grade 0 and grade 2 (alopecia) or grade 1 and grade 3 (erythema).

The metrics quantifying the interobserver variability are presented in table 4.11. All three metrics α_K, $\overline{\kappa_C}$, and *ICC* showed similar values. For alopecia, they ranged between 0.50 (α_K) and 0.58 (*ICC*) indicating a moderate agreement according to Landis and Koch (1977) and Koo and Li (2016). For erythema, the metrics ranged between 0.42 (*ICC*) and 0.49 (α_K) which is referred to as moderate (Landis and Koch, 1977) or even poor agreement (Koo and Li, 2016).

Interobserver variability with regard to different institutions

For each institute, the distribution of the severity grades for alopecia and erythema in terms of modal values for each of the 29 cases is given in table 4.10. For alopecia, the rate of grade 2 assessments was lowest for physicians from WPE (10.3 %) and highest at MGH (41.4 %). The highest rate of alopecia grade 0 was observed at UPTD (44.8 %). For erythema, no case was assessed with grade 3. In comparison to the overall modal value without any grade 2 occurrence, at least 3.4 % of the cases were rated with grade 2 in each institution, but not consistently the same patients. The highest rate of erythema grade 2 was reported by physicians from WPE and MGH (10.3 %).

Deviations of each institution's modal value from the overall modal value is shown in figure 4.12 for all cases. Exemplarily, for alopecia, physicians of WPE scored in 34 % of the cases lower and

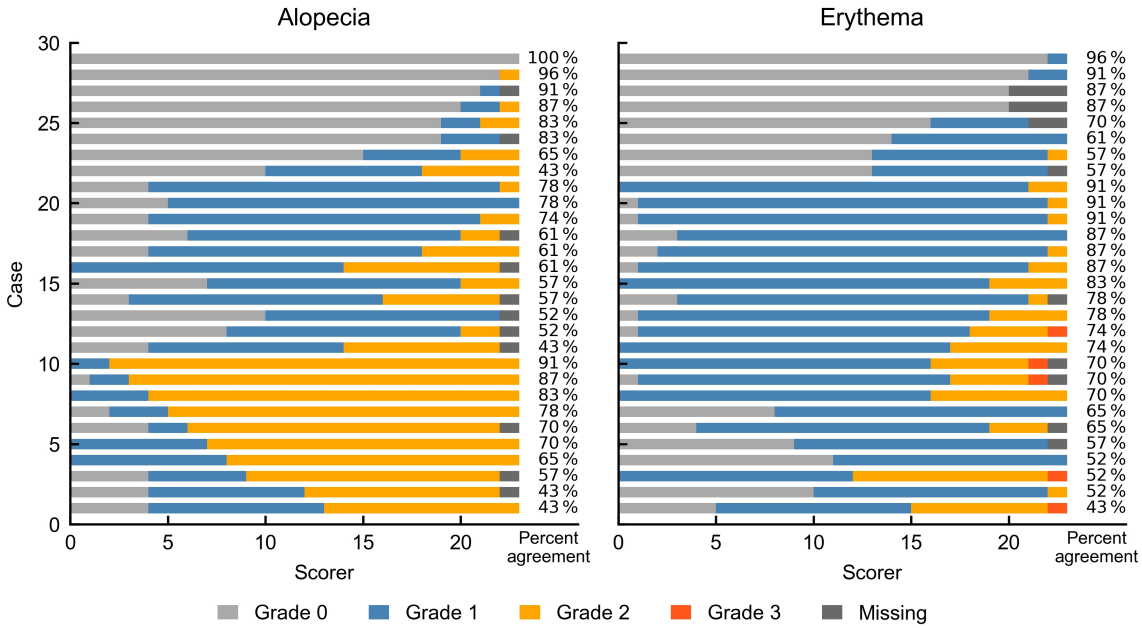


Figure 4.11: Distribution of the gradings for alopecia and erythema for all cases and all physicians sorted by severity grade (overall modal value) and percent agreement.

Table 4.11: Metrics assessing interobserver variability for the scoring of alopecia and erythema for all clincicians and at each institute separately.

	Alopecia						Erythema					
	All	WPE	MGH	UPTD	BER	TUEB	All	WPE	MGH	UPTD	BER	TUEB
ICC	0.58	0.37	0.69	0.48	0.61	0.65	0.42	0.36	0.49	0.45	0.45	0.61
α_K	0.50	0.39	0.68	0.41	0.61	0.62	0.49	0.43	0.56	0.50	0.51	0.63
$\overline{\kappa_C}$	0.51	0.38	0.69	0.45	0.60	0.64	0.45	0.39	0.51	0.47	0.50	0.61

Abbreviations: α_K, Krippendorff's alpha; *ICC*, intraclass correlation coefficient; $\overline{\kappa_C}$, average over pairwise Cohen's kappa; WPE, West German Proton Therapy Centre Essen; MGH, Massachusetts General Hospital in Boston; UPTD, University Proton Therapy Dresden; BER, Department of Radiation Oncology and Radiotherapy of the Charité Berlin; TUEB, Department of Radiation Oncology of the University Hospital Tuebingen.

in 3 % of the cases higher than the value determined by all physicians. At UPTD, the grading was lower than the overall grading in 21 % of the cases. At MGH, 10 % of the cases were rated higher than the overall grading by all physicians. The best agreement between institutional and overall modal values was observed among physicians from TUEB (93 %). For erythema, physicians at MGH rated 17 % of the cases higher than the modal value. At WPE, 14 % of the cases were rated higher and 3 % of the cases were rated lower than the overall modal values.

The metrics assessing the interobserver variability between physicians at each institute are given in table 4.11. In general, the values of the three individual metrics were similar for each institution. The agreement between physicians was generally slightly better for the assessment of alopecia than for erythema. The highest agreement was observed between physicians at MGH for assessing alopecia and between physicians from TUEB for assessing erythema (sub-

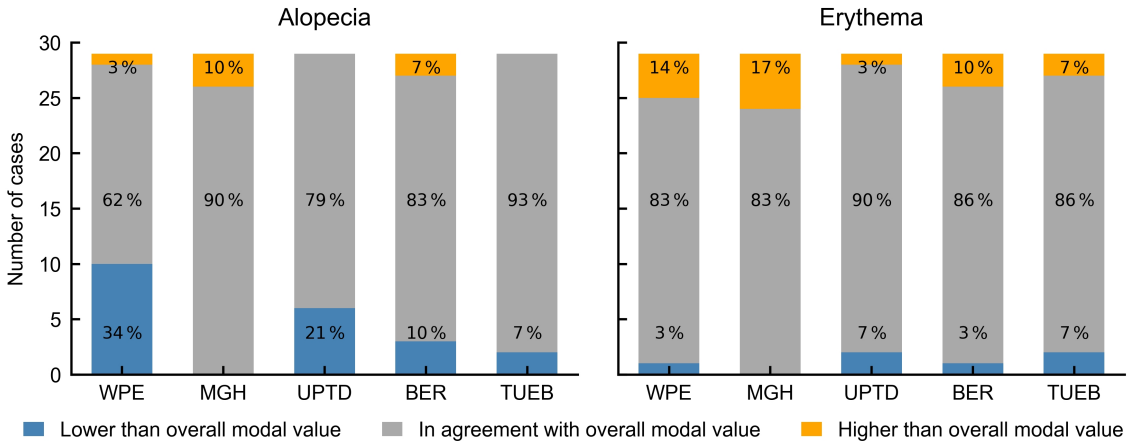


Figure 4.12: Deviations of the institutional modal values from overall modal values for the scoring of alopecia and erythema. WPE, West German Proton Therapy Centre Essen; MGH, Massachusetts General Hospital in Boston; UPTD, University Proton Therapy Dresden; BER, Department of Radiation Oncology and Radiotherapy of the Charité Berlin; TUEB, Department of Radiation Oncology of the University Hospital Tuebingen.

stantial/moderate agreement according to Landis and Koch (1977) and Koo and Li (2016), respectively). The lowest agreement was shown between physicians at WPE both for assessing alopecia and erythema (fair/poor according to Landis and Koch (1977) and Koo and Li (2016), respectively). Comparing the agreement between the institutes based on their individual modal values, a substantial/moderate agreement was observed for the assessment of alopecia (ICC = 0.72, α_K = 0.71, $\overline{\kappa_C}$ = 0.70). The agreement between institutes was moderate for erythema (ICC = 0.59, α_K = 0.62, $\overline{\kappa_C}$ = 0.58).

In order to reflect the real-life clinical situation, in which each case is assessed by only one physician from each institution, one physician's assessment was randomly drawn and the agreements between the five resulting assessments were investigated. The mean values for 1000 samples are presented in table 4.12 and were similar to those when comparing the agreement of all physicians, see table 4.11. The agreement for assessing alopecia and erythema was moderate and moderate/poor, respectively.

Table 4.12: Metrics of interobserver variability for assessing alopecia and erythema based on a random selection of one physician for each institute. The mean value of 1000 samples as well as the 95 % confidence interval is given as 2.5th and 97.5th percentile.

	ICC		α_K		$\overline{\kappa_C}$	
	Mean	**(95 % CI)**	**Mean**	**(95 % CI)**	**Mean**	**(95 % CI)**
Alopecia	0.55	(0.43 – 0.70)	0.53	(0.40 – 0.64)	0.54	(0.43 – 0.65)
Erythema	0.41	(0.26 – 0.58)	0.48	(0.35 – 0.61)	0.44	(0.31 – 0.57)

Abbreviations: α_K, Krippendorff's alpha; *ICC*, intraclass correlation coefficient; $\overline{\kappa_C}$, average over pairwise Cohen's kappa; CI, confidence interval.

4.4.3 Discussion

This study investigated to what extent physicians' assessments of erythema and alopecia differ when using the same grading system based on 23 radiation oncologists from five institutions. Furthermore, potential differences in grading between individual institutions were examined. Generally, the agreement between all physicians for assessing alopecia and erythema was moderate.

For the development of NTCP models, both the precise determination of dosimetric parameters and side effects are important. When discussing uncertainties of these input parameters, the focus laid mostly on the uncertainties of dose determination, for example, differences between planned and delivered dose (Shelley et al., 2017; McCulloch et al., 2018) and interobserver variabilities in contouring of OARs (Foppiano et al., 2003; Rosewall et al., 2011; Nelms et al., 2012b). Although the accuracy of the assessment of side effects may also impact the shape of the NTCP curves, uncertainties due to interobserver variability in scoring according to CTCAE have rarely been examined. Kapur et al. (2013) investigated the interobserver variability between 12 caregivers (6 radiation oncologists and 6 nurses) when assessing skin reactions of irradiated breast cancer patients based on images. When scoring according to CTCAE, a Fleiss-kappa value of 0.43 was revealed (moderate and poor agreement according to Landis and Koch (1977) and Koo and Li (2016), respectively). This value is comparable to the agreement for the grading of erythema in this thesis. The authors also examined the agreement when using an internally developed scoring system based on RTOG guidelines and literature reviews instead of the CTCAE scale. With this internal grading system, the agreement was much lower compared to CTCAE (Fleiss-kappa: 0.29). This clearly underlines that the use of standardised grading systems is essential, even if there is still potential for improvement. Based on the results by Kapur et al. (2013), Goyal et al. (2015) investigated whether variations between the severity grades for radiation dermatitis were related to free text toxicity assessment based on clinical records of 8 caregivers for 30 patient cases. As the agreement for terms that are explicitly included in the CTCAE scale showed comparatively low kappa values, the authors concluded that the assessment criteria were interpreted differently by the caregivers. In some cases, caregivers paraphrased the side effect with terms that indicate a certain severity grade, but they did not assign the corresponding but a lower grade. This implied that the examiners did not assign grades according to the exact definitions of the CTCAE, but according to other factors that do not necessarily correspond to the CTCAE terminology. Regular training sessions could help to improve the agreement between physicians in the assessment of side effects. It should be reviewed later on if there has been any real improvement.

The differences between the individual centres were not as distinct as it could be expected from the shift in the calibration for the model of acute alopecia, see section 4.2.2. There, lower incidence rates were observed in patients treated at WPE and MGH than predicted by the NTCP model. In this interobserver variability study, in 34 % of the cases evaluated by physicians at the WPE, lower values were assigned than the modal value of all physicians. At MGH, however,

slightly higher grades were scored than the modal value of all physicians. This effect could arise because the physicians participating in this interobserver variability study may not necessarily have been the same who examined the patients included in NTCP modelling. Since the agreement within the respective centres was also only moderate or fair, such variations may explain the differences in acute toxicity assessment.

The agreement between physicians was slightly better for scoring alopecia than erythema. This may be either because alopecia is easier to distinguish on an image or because the definitions of the CTCAE are more explicit. Moreover, alopecia is defined for only three different severity grades ($0-2$). Erythema, on the other hand, can be assessed with 6 different grades ($0-5$). For the cases included in this thesis, severity grades ranging from 0 to 3 were assigned. With a higher number of available grades, it is more difficult to achieve an exact agreement. Furthermore, not all severity grades were observed with similar frequency, e.g. erythema grade 3 was assigned in very few cases by few physicians, see figure 4.11. Krippendorff (2004) advised planning the reliability test such that all possible characteristics (severity grades) occur in sufficient variation, which is questionable for the assessment of erythema in this study.

The evaluation of the agreement between physicians is challenging, as there is no single *true* severity grade. Therefore, the modal value was chosen for each case in this study. Additionally, an agreement between physicians in the evaluation of side effects may also occur by chance alone. Therefore, a respective random distribution is shown in appendix B figure B.2. This distribution of percent agreement between 23 raters was generated assuming the same distribution of severity grades as in this study, but random drawing of grades from a normal distribution. The percent agreements of the physicians were above the 95^{th} percentile of the random distribution in 23/29 cases for alopecia and in 13/29 cases for erythema, which indicates a rather moderate agreement.

This study was limited by the small sample size of images and the restricted variation in the expression of different severity grades for erythema. Furthermore, the images are only surrogates and cannot reflect a real evaluation of the side effect during personal interaction with the patient, as it is performed in clinical practice. The physicians could not palpate, get a three-dimensional impression, examine in different lighting conditions or communicate with the patients. As a result, the agreement between the physicians in assessing these side effects may be higher in real-life.

In paper-based patient records, physicians may tick simultaneously two adjacent severity grades, which makes the evaluation considerably more complicated. Therefore, it is essential to implement a digital recording of side effects, which only allows for a clear assignment of one severity level. A further alternative could be the use of subjective QoL questionnaires, which do not require further interpretation by third persons and reflect the patient's condition from his point of view.

4.5 Summary

In this thesis, NTCP models for early and late side effects following cranial PBT have been developed and externally validated. Dose-volume parameters of the skin were associated to early alopecia and erythema. Validation confirmed that the presented dose-volume parameters are strongly associated with these endpoints for all investigated cohorts, but the calibration differed. Thus, centre-specific differences in toxicity assessment should be considered carefully. Late alopecia was also associated to dose-volume parameters of the skin while late hearing impairment was related to cochlear doses. Additionally, in analyses of the pooled cohort, dose-response relations between late physician-rated memory impairment and the dose to the remaining brain as well as the hippocampi. Late fatigue was related to dose to the brain stem and CTx. Due to the relatively low incidence rates and low severity of late side effects, NTCP models could only be developed for mild forms. In a model-based approach for patient selection for PBT, however, only models that predict side effects grade $\geq$ 2 would be considered (Langendijk et al., 2018). Although dose-response relationships for alopecia and dry eye syndrome grade $\geq$ 2 were found in GEE analyses, they still have to be validated.

To establish accurate dose-response relationships for side effects following modern RT techniques, multicentre studies are essential. Reliable NTCP models require high-quality toxicity data including extensive pre- and post-treatment examinations such as neurocognitive tests, audiograms or tests of endocrine function. Such tests are not subject to interobserver variability as with physician-rated side effects and are therefore more precise. But since they are considerably more time-consuming than a physician's evaluation, the most important side effects should be selected in new study designs. Prospective electronic recording of side effects (and their absence), comorbidities, medication and other possible cofactors is important to easily access and combine these data with imaging data and RT treatment parameters (Kessel et al., 2014).

In order to reduce the required time and manpower, a more comprehensive follow-up could possibly be achieved with larger follow-up intervals, e.g. biannually. Moreover, it should be ensured that all participating centres have the same follow-up interval. As there exist only a few PBT facilities in Germany, these have fairly large catchment areas, so that a lower frequency of follow-up may accommodate patients living in further distance. Other alternatives are study endpoints on QoL, as the patient can fill in questionnaires at home. Although these are highly subjective, they reflect the actual condition of the patient without interpretation by a physician and could be used for NTCP model endpoints.

5 Assessing the neurocognitive function following cranial proton beam therapy

Neurocognitive dysfunction is an important radiation-induced side effect as it is the most frequent complication among long-term survivors (Béhin and Delattre, 2003). Thus, several clinical trials on brain tumour patients include the assessment of neurocognitive function and QoL as relevant outcome measures (Taphoorn and Klein, 2004; Correa, 2010; Schagen et al., 2014). Both radiation dose and volume of healthy brain tissue exposed to radiation appear to be determining factors of neurocognitive function (Brown et al., 2003; Laack and Brown, 2004). Therefore, OARs and dosimetric parameters associated with changes in neurocognition have to be identified and considered for contouring and treatment planning in clinical routine. In particular, the irradiation of certain substructures of the brain such as the cerebellum (Merchant et al., 2014; Eekers et al., 2017) or the hippocampus (Gondi et al., 2012) may cause neurocognitive dysfunction. Besides several approaches in the development of advanced treatment planning techniques aiming to spare such brain structures (Gondi et al., 2014; Kazda et al., 2014; Tsai et al., 2015; Zhao et al., 2017), a potential dosimetric benefit of PBT has been discussed (Harrabi et al., 2016). However, it is still uncertain whether this dosimetric benefit leads to a potential reduction of radiation-induced neurocognitive dysfunction (Grosshans et al., 2017).

This exploratory study aims to investigate neurocognitive function assessed by the objective MoCA test and self-reported QoL in adults affected by brain tumours during a maximum follow-up of two years after primary or post-operative cranial PBT delivered with standard fractionation. Dose-volume parameters of several brain structures were investigated to find any relation to objective and subjective changes in neurocognition. Parts of the work presented in this chapter have been published (Dutz et al., 2020).

5.1 Patient cohort and experimental design

Patient data

The patient cohort analysed in this study is a subgroup of the cohort treated at UPTD described in chapter 3. It included patients diagnosed with LGG, HGG, meningioma, pituitary adenoma or craniopharyngioma that were consecutively treated with PBT between 07/2015 and 07/2017 within the prospective clinical trial (DRKS00007670, 2015). The trial comprised 90 patients at the time of data analysis (02/2019). For this study, 62 patients were selected who met further criteria: normo-fractionated PBT and available baseline data on neurocognitive function, see figure 5.1. The patient cohort is characterised in table 5.1. This retrospective study was approved by the local Ethics Committee (EK48012019).

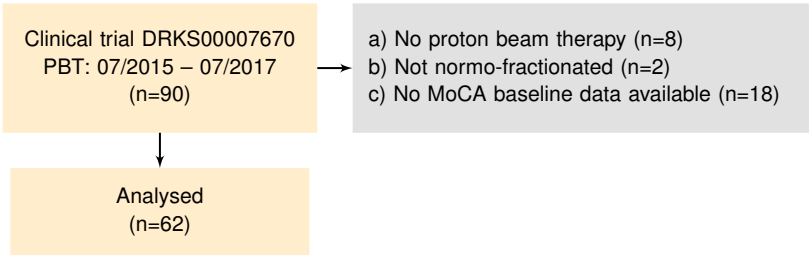


Figure 5.1: Design of the study on neurocognition. Adapted from Dutz et al. (2020).

Radiotherapy planning and treatment delivery

The planning CT scans were rigidly registered with (post-operative) T2- and T1-weighted MRI scans with contrast agent applied. GTVs encompassed the macroscopic tumour or the resection cavity including the residual tumour, if present. CTVs included microscopic disease and oedema and were created considering tumour histology as well as anatomical boundaries. Treatment plans were calculated using a DS PBT technique (XiO®). For treatment planning, a constant RBE of 1.1 was anticipated. Dose prescription ranged from 30 to 60 Gy(RBE) with 2 Gy(RBE) per fraction. Patient positioning was verified daily by manual registration of orthogonally paired X-ray images with digitally reconstructed radiographs of the planning CT. Treatment planning is described in detail in section 4.1.

Endpoint definition

Neurocognitive function and QoL were assessed at baseline prior to PBT on the day of the planning CT scan. Regular clinical follow-up visits were performed three months after treatment and at intervals of three months thereafter until disease progression occurred. The MoCA test was conducted by trained study nurses. No significant differences in MoCA total scores between different interviewers were observed. Neurocognitive function was assessed by the German version of the paper-and-pencil MoCA test (Nasreddine et al., 2005), see section 2.3.3. To minimise learning effects, three alternative test versions were iterated between each assessment (Cooley et al., 2015). At baseline, all patients started with the same test version. QoL was assessed by the validated EORTC QLQ-C30 version 3.0 (Aaronson et al., 1993) and the brain-tumour specific module EORTC QLQ-BN20 (Osoba et al., 1996) in German. The evaluation of QoL was restricted to relevant scales in the context of PBT of brain tumour patients including global health status, cognitive, physical, emotional, social and role function, fatigue, pain and insomnia (EORTC QLQ-C30). The items future uncertainty, headache, seizures, hair loss, visual disorder and communication deficit were selected from EORTC QLQ-BN20. QLQs were evaluated according to the EORTC scoring guidelines (Fayers et al., 2001), see section 2.3.2. To investigate the association between the objective MoCA test and self-reported QoL, subitems from both measures were correlated: QLQ cognitive function (EORTC QLQ-C30) with MoCA total

Table 5.1: Detailed patient characteristics of the study on neurocognition.

Characteristic	**Median**	**Range**
Age at PBT in years	47.0	18.8 – 84.7
Tumour volume (CTV) in cm^3	119.9	1.2 – 472.8
Prescribed dose in Gy(RBE)	57.0	30.0 – 60.0
Recurrence-free follow-up in months	22.5	0.0 – 37.2
	N	**(%)**
Gender [female/male]	23/39	(37.1/62.9)
Chemotherapy [yes/no]	24/38	(38.7/61.3)
Surgery [yes/no]	53/9	(85.5/14.5)
Corticosteroids at baseline [yes/no]	25/33	(40.3/53.2)
Impairment at baseline [MoCA$<$26/MoCA$\geq$26]	30/32	(48.4/51.6)
Re-irradaition [yes/no]	15/47	(24.2/75.8)
Tumour histology		
Glioma II	4	(6.5)
Glioma III	16	(25.8)
Glioma IV	20	(32.2)
Meningioma	8	(12.9)
Other	14	(22.6)
IDH1 [mutant/wild-type/missing]		
Glioma II	3/1/0	(75.0/25.0/0.0)
Glioma III	11/4/1	(15.0/75.0/10.0)
Glioma IV	3/15/2	(68.7/25.0/6.3)
MGMT [unmethylated/methylated/missing]		
Glioma II	0/1/3	(0.0/25.0/75.0)
Glioma III	4/8/4	(25.0/50.0/25.0)
Glioma IV	11/9/0	(55.0/45.0/0.0)
Tumour location* [left/right/central]	21/29/11	(33.9/46.8/17.7)
Frontal lobe	1/8/1	(1.6/12.9/1.6)
Temporal lobe	12/12/1	(19.4/19.4/1.6)
Parietal lobe	5/1/0	(8.1/1.6/0.0)
$>$1 lobe	2/4/0	(3.2/6.5/0.0)
Other lobe	0/2/9	(0.0/3.2/14.5)
ECOG Performance Status [0/1/2]	34/24/4	(54.8/38.7/6.5)
Local recurrence after PBT	21	(27.4)

Abbreviations: IDH, isocitrate dehydrogenase; MGMT, methylguanine-DNA methyltransferase; ECOG, Eastern Co-operative of Oncology Group; *One patient had a lesion in the left temporal lobe and another lesion in the right frontal lobe.

score, QLQ communication deficit (EORTC QLQ-BN20) with MoCA language score, QLQ visual disorder (EORTC QLQ-BN20) with MoCA visuospatial examination score and single item QLQ memory (EORTC QLQ-C30, 4-point scale) with MoCA memory score.

The impact of the following clinical cofactors on test results was investigated: age at PBT, gender, ECOG status, tumour histology, location of the tumour (side and lobe), tumour volume (CTV), surgery, re-irradiation, CTx (TMZ), administration of corticosteroids (dexamethasone) and impaired MoCA total score at baseline ($<$ 26).

Extraction of dose-volume parameters

Different brain structures were retrospectively contoured on registered MRI scans by experienced radiation oncologists. These included hippocampi, amygdalae, thalami, cerebellum (pos-

terior, anterior and entire organ), frontal and temporal lobes using specific contouring guidelines (Schmahmann et al., 1999; Chera et al., 2009; Sun et al., 2014; Scoccianti et al., 2015; Eekers et al., 2018). All dose distributions were retrospectively exported using RayStation® scripts. Dose-volume parameters were extracted for the above-mentioned structures as well as the entire brain and brain excluding the CTV (brain-CTV). Relative volume parameters *VxGy*(*RBE*) ranging from 10 Gy(RBE) to 60 Gy(RBE) in increments of 10 Gy(RBE), mean and median doses and the near-maximum dose *D*2% were analysed.

Statistical analyses

Differences in MoCA scores between two patient groups classified according to binary clinical cofactors were evaluated by Mann-Whitney U tests. If a binary cofactor-based classification was not applicable, the Kruskal-Wallis test has been applied. Correlations between continuous variables were assessed using the Spearman correlation coefficient ρ. Analyses were restricted to a maximum follow-up period of 24 months as only a few patients ($n < 20$) remained for evaluation after this time point (response rates are given in appendix C table C.1). Test scores between follow-up and baseline were compared using the Wilcoxon signed-rank test. Changes in QoL were scaled to *no change*, *a little*, *moderate* and *very much* (Osoba et al., 1998), see section 2.3.2. For the MoCA test, differences of 4 or more points were defined as *large*, since the threshold for impairment is defined as 26 out of 30 points by the authors of the MoCA test (Nasreddine et al., 2005).

To investigate differences from baseline over time, repeated-measure analysis (mixed-model analysis) with first-order autoregressive covariance structure and restricted maximum likelihood was conducted. This included clinical cofactors, time after PBT and the corresponding interaction term between time and the cofactor as fixed-effects covariates. For MoCA and QLQ cognitive function, dose-volume parameters were also included as predictors in univariable models. All dose-volume parameters showing a significant association with differences of the MoCA score or QLQ cognitive function compared to baseline were included in multivariable models with clinical cofactors that also showed a significant association in univariable analysis. Mixed-model analyses were also performed on all subscales of the MoCA test.

For visualisation, patients were classified into two groups according to the dosimetric parameters that were significant in mixed-model analyses. To define the classification threshold, the dosimetric parameters were dichotomised at several cut-off values. Mixed-model analyses were performed including time, dichotomised dosimetric parameter and their interaction. The dichotomised dosimetric parameter with the lowest p-value was selected as the classification threshold. For these two patient groups, changes in neurocognitive function compared to baseline were visualised over time.

For all mentioned statistical analyses, two-sided tests were performed using IBM SPSS Statistics 25 and in-house written Python (version 2.7.10) programmes using the module scipy (statsmodels). P-values < 0.05 were considered statistically significant.

5.2 Results

5.2.1 Performance at baseline

At baseline, the mean MoCA total score was 24.8 with a standard deviation (SD) of 4.2, indicating a minor cognitive impairment according to the threshold of 26 defined by Nasreddine et al. (2005), see table 5.2. Significantly lower MoCA scores were observed for patients with worse (higher) ECOG performance status (p = 0.036) and tumour location in the left hemisphere (p = 0.009). Grade II glioma patients performed significantly better than patients with different histology (p = 0.043). Self-reported QLQ cognitive function was 68.9 (SD = 26.6) at baseline indicating a slightly decreased performance compared to healthy populations (Michelson et al., 2000). High self-reported cognitive function was associated with good ECOG performance status at baseline (p = 0.006) or prescription of concomitant CTx (p = 0.021). For other QoL items from the QLQ-BN20 and QLQ-C30 questionnaires, mean baseline values are given in appendix C table C.1.

Patients who were classified as impaired according to their MoCA test result (MoCA total score $<$ 26) reported lower cognitive function compared to normal performing patients (p = 0.028) at baseline. These patients also described significantly lower global health status, physical and role function as well as more fatigue, pain, headache and communication deficits compared to normal performing patients (MoCA total score $\geq$ 26) at baseline. A tumour location in the left hemisphere was significantly associated with lower global health status, worse emotional and role function, a higher level of fatigue and communication deficits. A high ECOG performance status was associated with a reduced condition of a variety of items: global health status, social, emotional, physical and role function as well as higher levels of fatigue, pain and headache. CTx was associated with higher role function and lower fatigue, headache and fewer communication deficits at baseline. Women complained significantly more about pain, headache and hair loss compared to men. Re-irradiated patients had significantly higher fatigue and lower role function than others.

5.2.2 Correlation between subjective and objective measures

At most time points, MoCA total score and QLQ cognitive function, MoCA memory score and QLQ memory as well as MoCA language score and QLQ communication deficit were significantly correlated, see table 5.3. Especially at later follow-up visits ($>$ 18 months), these items were highly correlated ($\rho > 0.70$, $p < 0.001$). However, no or only a low correlation was observed for MoCA visuospatial examination score and QLQ visual disorder.

Table 5.2: Comparison of the MoCA total scores and self-reported cognitive function according to the EORTC-QLQ-C30 questionnaire (QLQ cognitive function) at baseline between patient groups.

	MoCA test		**QLQ cognitive function**	
	Mean	**p-value**	**Mean**	**p-value**
Entire cohort	24.8		68.9	
Gender [female/male]	25.0/24.7	0.54	70.5/67.9	0.85
Surgery [yes/no]	24.7/25.3	0.32	67.6/75.9	0.31
Chemotherapy [yes/no]	23.8/25.4	0.075	77.1/63.5	0.021
Corticosteroids [yes/no]	23.9/25.9	0.079	61.8/71.7	0.29
Re-irradiation [yes/no]	25.2/24.7	0.68	65.5/69.9	0.43
Impairment at baseline [MoCA$<$ 26/MoCA$\geq$ 26]	21.3/28.1	$<$ 0.001	60.6/76.9	0.028
Tumour histology		0.043	79.2	0.47
Glioma II	29.5		79.2	
Glioma III	23.9		67.8	
Glioma IV	23.9		71.7	
Meningioma	25.3		56.3	
Other	25.6		70.2	
Tumour location		0.15		0.86
Frontal	26.2		70.0	
Temporal	23.8		69.3	
Parietal	22.2		70.0	
Occipital	24.0		44.4	
$>$1 lobe	25.5		75.0	
Other	26.8		69.4	
Tumour location [left/right/central]	22.3/25.8/26.6	0.009	59.5/75.6/69.7	0.084
ECOG performance status[0/1/2]	25.9/24.0/20.0	0.036	77.3/62.5/37.5	0.006
	Spearman ρ	**p-value**	**Spearman ρ**	**p-value**
Tumour volume (CTV) in cm^3	-0.14	0.27	-0.05	0.71
Age at PBT in years	-0.18	0.16	0.16	0.22
Prescribed dose in Gy(RBE)	0.04	0.74	0.14	0.27

Abbreviations: CTV, clinical target volume; ECOG, Eastern Co-operative of Oncology Group.

Table 5.3: Spearman correlation (correlation coefficient ρ) between MoCA items and self-reported quality of life items at different time points after PBT.

			Baseline	**3 months**	**6 months**	**9 months**	**12 months**	**15 months**	**18 months**	**21 months**	**24 months**
MoCA item	**QLQ item**	**N**	61	48	37	35	27	27	21	24	26
Total score$^+$	Cognitive function$^+$	ρ	0.29	0.46	0.59	0.49	0.29	0.38	0.72	0.76	0.71
		p-value	0.023	0.001	$<$ 0.001	0.003	0.14	0.053	0.001	$<$ 0.001	$<$ 0.001
Memory$^+$	Memory$^-$	ρ	-0.17	-0.43	-0.26	-0.45	-0.38	-0.37	-0.51	-0.79	-0.70
		p-value	0.2	0.002	0.13	0.008	0.05	0.066	0.018	$<$ 0.001	$<$ 0.001
Language$^+$	Communication deficit$^-$	ρ	-0.52	-0.31	-0.55	-0.41	-0.24	-0.12	-0.69	-0.77	-0.48
		p-value	$<$ 0.001	0.03	$<$ 0.001	0.016	0.24	0.56	0.001	$<$ 0.001	0.020
Visuospatial examination$^+$	Visual disorder$^-$	ρ	0.02	-0.09	-0.36	-0.11	-0.06	-0.41	-0.43	-0.47	-0.21
		p-value	0.88	0.53	0.027	0.52	0.78	0.036	0.051	0.021	0.35

Abbreviation: N, number of patients. $^+$ high score represents high function, $^-$ high score represents low function

5.2.3 Time-dependent score analyses

The median recurrence-free follow-up time was 22.5 months, see table 5.1. The mean MoCA total score and QLQ cognitive function remained stable over time, see figure 5.2. No statistically significant difference was observed between baseline values and test results at any recurrence-free time after PBT for both measures. The individual changes in both MoCA total score and QLQ cognitive function are depicted in figure 5.3. The majority of patients (n = 29) experienced no or only small changes in MoCA score averaged over 6 to 24 months after PBT. At the same time, two patients improved their test result by $\geq$ 4 points and seven patients by 2 to 4 points. Four patients deteriorated by 2 to 4 points and 4 patients by more than 4 points. The mean self-reported cognitive function of the entire cohort showed no or small changes. Averaged over 6 to 24 months after PBT, twelve patients reported a moderate (n = 7) or even large (n = 5) improvement in QLQ cognitive function. However, twelve patients deteriorated to a large extent.

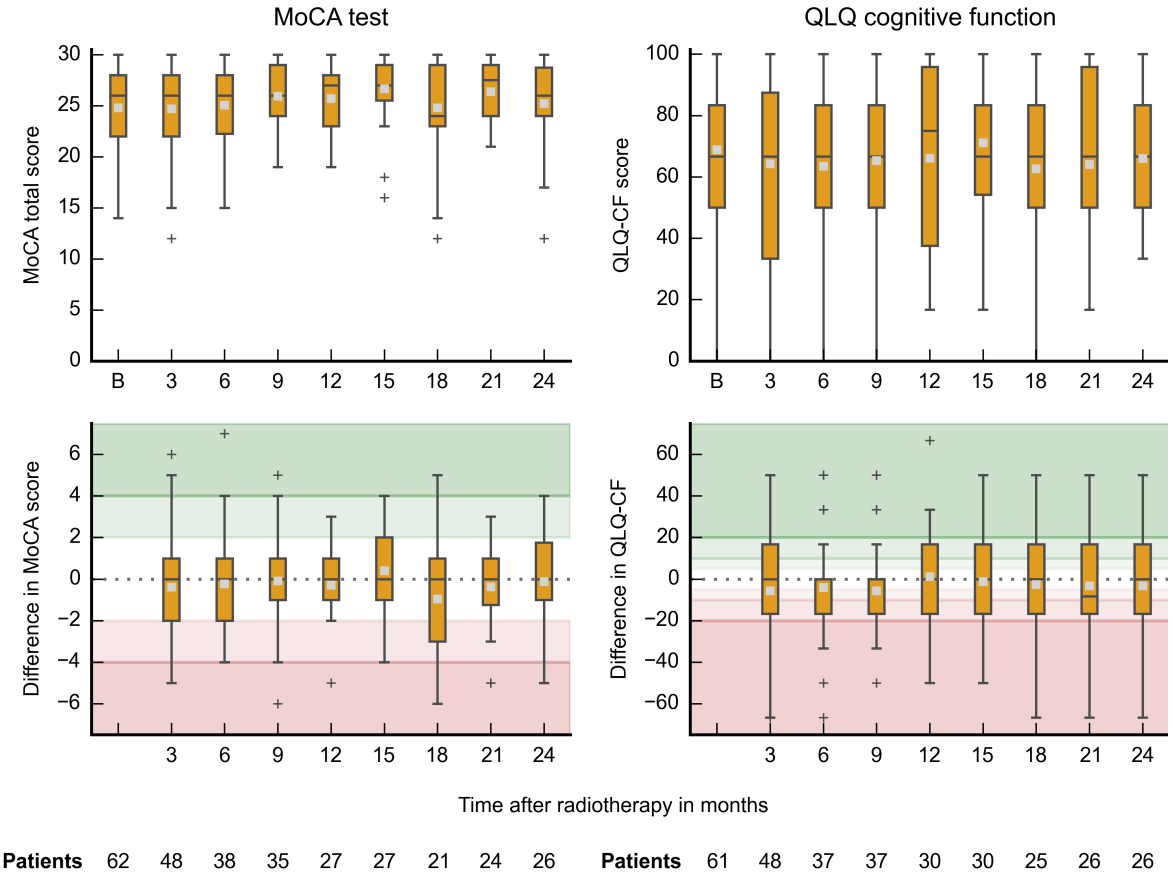


Figure 5.2: MoCA total score and self-reported cognitive function (QLQ cognitive function) over time. Absolute scores are shown in the top row, differences to baseline (B) below. Shaded areas mark the extent of changes in QLQ and MoCA (red: worsening, green: improvement). Extent of QLQ changes from light to dark: a little (5.0 – 9.9), moderate (10.0 – 19.9) and very much ($\geq$ 20.0) (Osoba et al., 1998). Extent of MoCA changes from light to dark: moderate (2 – 4 points), large ($\geq$ 4 points). Adapted from Dutz et al. (2020).

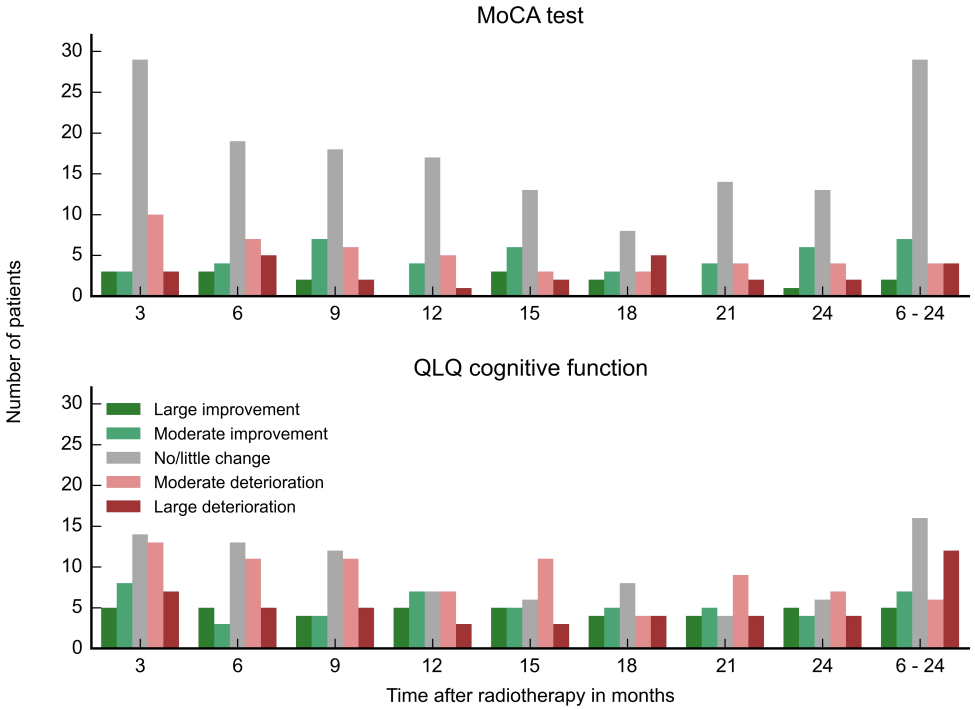


Figure 5.3: Clinically relevant changes in MoCA scores (top) and self-reported (quality of life questionnaire, QLQ) cognitive function (bottom). Last column: mean change between 6 to 24 months. The extent of changes for QLQ cognitive function as defined by Osoba et al. (1998) and for the MoCA test as follows: moderate (2 to 4 points), large (4 or more points). Adapted from Dutz et al. (2020).

In mixed-model analyses, MoCA differences compared to baseline were associated with tumour location: patients with a tumour location in the left hemisphere improved significantly more compared to patients with central tumours (p = 0.036), see table 5.4. No clinical cofactor was significantly associated with changes in self-reported cognitive function compared to baseline. For other selected QoL items, results of mixed-model analyses are given in appendix C table C.2.

The mixed-model analyses including dose-volume parameters of different brain structures revealed a significantly reduced MoCA score with increasing $V30Gy(RBE)$ and $V40Gy(RBE)$ of the anterior cerebellum, see table 5.4. Classification thresholds were $V30Gy(RBE) \geq 35\,\%$ and $V40Gy(RBE) \geq 20\,\%$. Figure 5.4 shows MoCA differences over time for patient groups classified according to these thresholds.

In multivariable mixed-model analyses including $V30Gy(RBE)$ or $V40Gy(RBE)$ of the anterior cerebellum and tumour location (left/right hemisphere or central), both dose-volume parameters of the anterior cerebellum were still significantly associated with a reduced MoCA score compared to baseline ($p < 0.010$), see table 5.4. Both dosimetric parameters $V30Gy(RBE)$ and $V40Gy(RBE)$ of the anterior cerebellum were highly correlated (ρ = 0.96, $p < 0.001$). Both time and the interaction between time and dose-volume parameters were not significantly associated with the MoCA score differences compared to baseline. This indicates that the effect of the dosi-

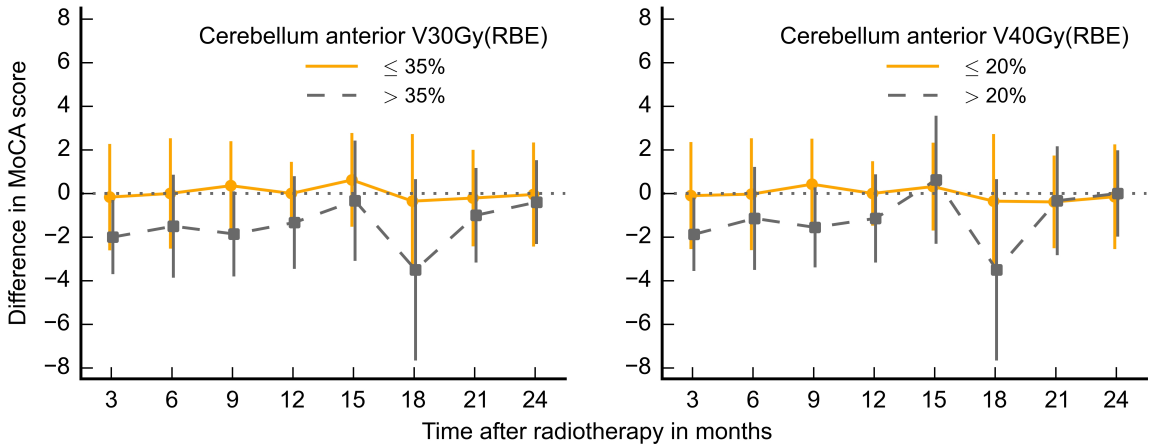
Table 5.4: Results of the mixed-model analyses on differences of the MoCA score to baseline including all time points. Model coefficients and standard deviation of the clinical or dosimetric parameters, time after proton beam therapy and their interaction term are given. Presented are univariable (top) and multivariable (bottom) models with significant associations.

Model parameter	Coefficient	SD	p-value
Univariable analysis			
Location [reference: central]			
Left	2.31	1.09	0.036
Right	1.24	1.03	0.24
Time after treatment	0.17	0.17	0.30
Interaction of time and location [reference: central]			
Left	-0.19	0.22	0.38
Right	-0.21	0.21	0.31
Constant	-1.75	0.86	
Cerebellum anterior $V30Gy(RBE)$	-4.28	2.08	0.042
Time after treatment	-0.03	0.09	0.77
Interaction of time and cerebellum anterior $V30Gy(RBE)$	0.43	0.41	0.30
Constant	0.11	0.46	
Cerebellum anterior $V40Gy(RBE)$	-5.63	2.69	0.039
Time after treatment	-0.03	0.09	0.73
Interaction of time and cerebellum anterior $V40Gy(RBE)$	0.61	0.54	0.27
Constant	0.11	0.45	
Multivariable analysis			
Location [reference: central]			
Left	2.37	1.00	0.020
Right	1.34	0.95	0.16
Cerebellum anterior $V30Gy(RBE)$	-5.14	1.93	0.009
Time after treatment	0.11	0.16	0.50
Interaction of time and location [reference: central]			
Left	-0.18	0.20	0.36
Right	-0.18	0.19	0.34
Interaction of time and cerebellum anterior $V30Gy(RBE)$	0.58	0.38	0.13
Constant	-1.16	0.79	
Location [reference: central]			
Left	2.42	0.99	0.018
Right	1.42	0.95	0.14
Cerebellum anterior $V40Gy(RBE)$	-6.85	2.51	0.008
Time after treatment	0.10	0.15	0.50
Interaction of time and location [reference: central]			
Left	-0.19	0.20	0.35
Right	-0.19	0.19	0.31
Interaction of time and cerebellum anterior $V40Gy(RBE)$	0.87	0.51	0.090
Constant	-1.21	0.78	

Abbreviation: SD, standard deviation.

metric parameters on the MoCA change compared to baseline did not change over time. In this study, no association between changes in the MoCA total score and dose-volume parameters of further brain structures such as hippocampus were observed, see appendix C table C.3.

Results of the mixed-model analyses for the MoCA subscales are given in appendix C table C.4. Deterioration in the MoCA memory domain over time was significantly associated with most dosimetric parameters of the anterior cerebellum as well as $D2\%$ of the entire cerebellum.



Patients									Patients								
≤ 35%	42	32	28	21	21	17	19	21	≤ 20%	40	31	26	20	19	17	18	20
> 35%	6	6	7	6	6	4	5	5	> 20%	8	7	9	7	8	4	6	6

Figure 5.4: Mean MoCA differences from baseline over time for patient subgroups classified according to the volume of the anterior cerebellum receiving 30 Gy(RBE) and 40 Gy(RBE). Error bars represent standard deviations. Adapted from Dutz et al. (2020).

The association of MoCA memory score with $V40Gy(RBE)$ of the anterior cerebellum revealed the lowest p-value (p = 0.005). No significant relations between deterioration in MoCA memory and doses to the hippocampi or temporal lobes could be observed. No associations were found between dose-volume parameters and self-reported cognitive function.

5.3 Discussion and conclusion

In this study on brain tumour patients treated with PBT, self-reported and objectively measured neurocognition and many other considered QLQ domains remained largely stable over time during recurrence-free follow-up. The objective and self-reported measures for neurocognition were highly correlated. Patients with impaired neurocognition on the MoCA test at baseline performed worse on many QLQ items compared to normal performing patients. An higher relative volume of the anterior cerebellum exposed to doses of 30 to 40 Gy(RBE) was associated with a slight deterioration of the MoCA score.

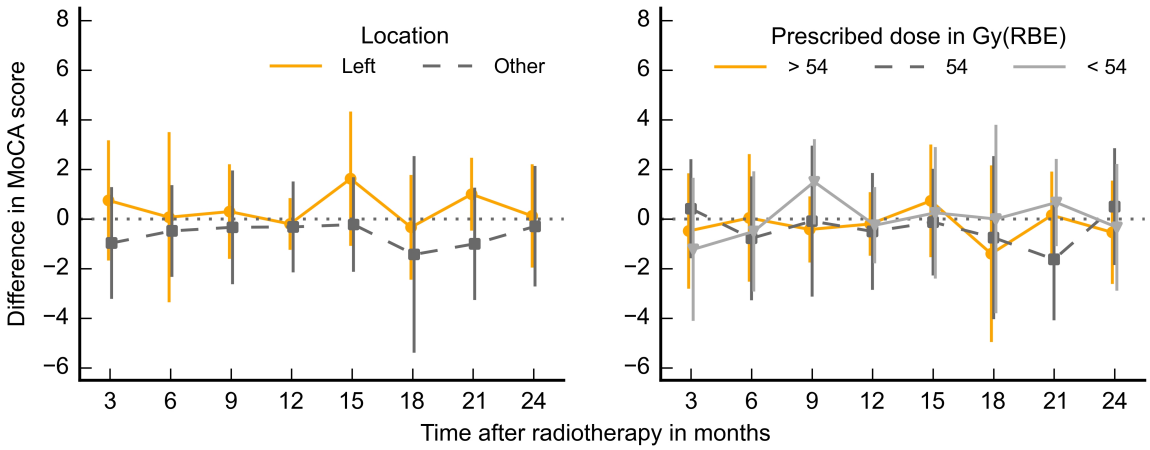
Prior to treatment, the mean MoCA total score was similar to that of other brain tumour patients (Robinson et al., 2015; Schiavolin et al., 2018) and healthy populations (Rossetti et al., 2011; Thomann et al., 2018), see appendix C table C.5. Among the investigated QoL items at baseline, patients in this study reported lower functional QoL and higher fatigue compared to healthy individuals (Michelson et al., 2000; Schwarz and Hinz, 2001; Scott et al., 2008), but similar values compared to other reported brain tumour patients (Scott et al., 2008; Renovanz et al., 2018). Symptoms of the brain-tumour specific questionnaire QLQ-BN20 in this study were similar compared to other brain tumour patients (Taphoorn et al., 2010; Renovanz et al., 2018).

Neurocognitive dysfunction before any treatment in diffuse glioma patients has been described previously (van Kessel et al., 2017). However, studies reporting the cognitive outcome of adult patients treated with protons are rare. Two studies based on the same small cohort of 20 adult LGG patients treated with PBT did not find any evidence for an overall decline in neurocognitive function or QoL during 5-year follow-up (Shih et al., 2015; Sherman et al., 2016). Test performance did not differ significantly between LGG patients and normative data (Sherman et al., 2016).

In this study, most items from the objective MoCA test correlated to self-reported neurocognitive function. Several studies investigating the association between objective and self-reported measures of neurocognition reported only moderate correlations (Klein et al., 2002; Li et al., 2008; Olson et al., 2011a; Boele et al., 2014). Olson et al. (2011a) observed a moderate correlation of the MoCA test with self-reported community integration (Spearman ρ = 0.35) that assesses disability and functioning in home, social interactions and productive activities. The function domain of the self-reported brain-tumour specific QLQ FACT brain tumour module (FACT-Br) measuring the ability to work, sleep, enjoy life and regular activities was associated with the visuospatial-executive domain of the MoCA test (Spearman ρ = 0.37) (Olson et al., 2011a). For the same cohort, Boele et al. (2014) observed weak to moderate correlations between the self-reported physical function of the 36-item Short-Form Health Survey (SF-36) and QLQ-BN20 with many neurocognitive domains measured with different neuropsychological tests. The correlation of MoCA test results with QoL items at baseline is in line with a study conducted by Noll et al. (2017) that analysed 103 glioma patients observing a relation between impaired cognition and reductions in QoL even prior to treatment. The high correlation of both measures in the present study showed that the MoCA test reflects the neurocognitive condition of the patients quite well.

Most QoL scales showed no or only minor changes during follow-up in this study in contrast to other studies on patients treated with cranial XRT (Bosma et al., 2007; Aaronson et al., 2011; Cole et al., 2013). Aaronson et al. (2011) described a decreased performance in many QoL items after 5.6 years from diagnosis in 195 LGG patients treated with or without XRT. Self-reported neurocognitive dysfunction among these patients was associated with reduced QoL. In contrast to most studies describing the course of neurocognition in LGG patients, Bosma et al. (2007) applied standardised tests in HGG patients at 8 months (n = 14) and 16 months (n = 18) following XRT. During this time, neurocognitive function decreased considerably and patients with tumour progression showed even worse results than patients without progression.

In the present study, MoCA differences compared to baseline were associated with the lateral location of the tumour. Although patients with left-sided tumours performed significantly worse on the MoCA test at baseline, they improved more after PBT compared to patients with a right-sided or central tumour, see figure 5.5. This may be due to the location of brain areas responsible for neurocognitive processes (lateralisation), e.g. language processing is mainly left lateralised (Frost et al., 1999). Greater impairment in verbally mediated domains in patients with left-sided compared to right-sided tumours has previously been described (Surma-aho et al., 2001; Klein



Patients								
Left	16	14	10	5	8	6	7	8
Other	31	23	24	22	18	14	16	17

Patients								
> 54	27	23	19	15	15	10	13	13
54	12	9	12	8	8	8	8	10
< 54	9	6	4	4	4	3	3	3

Figure 5.5: MoCA difference from baseline over time, patient classification according to tumour location in the left hemisphere vs other location (right hemisphere or central location) and prescribed dose. Error bars represent standard deviations. Adapted from Dutz et al. (2020).

et al., 2002; Schagen et al., 2014). For patients with left-sided LGG, Sherman et al. (2016) observed a significant impairment on verbal measures prior to PBT and an improvement after PBT. Raysi Dehcordi et al. (2013) showed improvement after surgery, while Correa et al. (2007) did not find any difference in neurocognitive function according to tumour location after XRT.

Deterioration of the MoCA score was associated with an increased relative volume of the irradiated anterior cerebellum exposed to doses of 30 to 40 Gy(RBE) in this study. Deficits on several neurocognitive domains as well as acute fatigue have previously been described as a result of cerebellar damage (Schmahmann, 2010; Grimaldi and Manto, 2012; Gulliford et al., 2012a; Merchant et al., 2014; Eekers et al., 2017). In a study on 76 children with infratentorial ependymoma treated with adjuvant XRT up to 59.4 Gy, Merchant et al. (2014) observed correlations of dose to the posterior cerebellum and deficits in intelligence quotient during 5-year follow-up. Furthermore, they found significant associations of declined intelligence quotient and mean dose to infratentorial as well as anterior and posterior cerebellar volumes. A decline in academic achievement, as another measure of cognition, was mainly related to dose to the posterior cerebellum.

The present study comprised re-irradiated patients that may show a different neurocognitive performance over time compared to primary brain tumour patients. However, in multivariable mixed-model analyses including re-irradiation as a separate parameter, deterioration of the MoCA score was still significantly associated with increased $V30\,Gy(RBE)$ and $V40\,Gy(RBE)$ of the anterior cerebellum. Even in analyses excluding re-irradiated patients, these dosimetric parameters were significantly associated with neurocognitive decline. In order to spare the cerebellar regions from radiation dose, the cerebellum may be considered as OAR and therefore routinely delin-

eated. However, due to the relatively small sample size of our analysis, these findings should be validated in larger studies.

Against the expectations, no relation between neurocognitive decline and the prescription dose could be shown, see figure 5.5. Furthermore, no relation between neurocognitive dysfunction and irradiation of the hippocampi could be observed as previously hypothesised (Gondi et al., 2012), see appendix C table C.3. Apart from the issue that hippocampal dose-volume parameters could not be determined in patients with a tumour or resection cavity in the hippocampi region, it is furthermore possible that minor neurocognitive changes cannot be detected accurately with the brief MoCA test. It may not be sensitive enough to detect impairment in every neurocognitive domain (Chan et al., 2014; Robinson et al., 2015). Results of subscales of the MoCA test are combined into a total score, whereby information about individual neurocognitive domains may be lost. Hence, the subscales were analysed separately, showing associations between cerebellar dose and deterioration in the MoCA memory subscale. Furthermore, differences to baseline of the MoCA total score were correlated to differences of the MoCA memory subscale (Spearman $\rho = 0.68$). Thus, it is consistent that similar dosimetric parameters of the cerebellum were related to the deterioration of both the MoCA total score and the MoCA memory score in this analysis.

To standardise neurocognitive assessment, comprehensive series of tests have been recommended for individualised neuropsychological assessment of both LGG and HGG patients. These testing batteries include ten different tests on memory, attention and executive function, intellectual function and information processing (Taphoorn and Klein, 2004). However, these extensive neuropsychological tests are time-consuming, require substantial manpower and may cause fatigue and reduced compliance of brain tumour patients. For clinical studies with continuous follow-up, brief (ten minutes) tests are more feasible (Taphoorn and Klein, 2004; Meyers and Brown, 2006). Among these tests, the MoCA test showed superior sensitivity in detecting neurocognitive impairment and correlation to QoL compared to the widely used MMSE (Olson et al., 2011a; Olson et al., 2011b). In future studies, neurocognitive tests could be performed less frequently, but with an extended test battery that could provide more detailed information about different neurocognitive domains. A longer time interval between interviews could further reduce possible learning effects, although three different test versions have already been used in this study. In an even more detailed follow-up, more specific questionnaires could be used that only address the subjective perception of neurocognitive function such as the FACT-Cog questionnaire (Wagner et al., 2009). The comprehensive EORTC-QLQs include only a few items concerning self-reported neurocognitive function.

The detection of correlations with dosimetric parameters is challenging as neurocognitive sequelae can be consequences of multiple causes: the tumour itself (Klein et al., 2002; Klein et al., 2003), treatment with neurosurgery, RT, CTx, antiepileptics or corticosteroids or comorbidities. Additionally, patient-related factors such as tumour regrowth, leptomeningeal involvement, metabolic disturbances or tumour-related epilepsy may have an impact (Taphoorn and Klein, 2004).

Limitations A limitation of this study is the heterogeneity of the rather small cohort with regard to different tumour histologies. While patients with benign tumours or LGGs often show a rather long survival, patients with HGGs, especially glioblastomas and IDH wild-type gliomas, tend to have poorer survival. Inclusion of HGG patients may lead to bias in the evaluation of the time course of neurocognitive function. However, all four patients with IDH1 wild-type grade III glioma had a survival of more than 29 months and 68 % of all glioblastoma patients were still alive 12 months after PBT. Patients who were excluded from follow-up due to death or local recurrence did not perform significantly worse at their last valid follow-up visit than patients with local control, see appendix C figure C.1. Despite a relatively long median follow-up period of 22.5 months, an even longer follow-up would be desirable, since neurocognitive deterioration associated with dose to the brain may manifest many years after treatment (Douw et al., 2009).

For treatment planning, RBE was considered to be constant at 1.1. If the distal edge of the Bragg peak with potentially increased RBE (Lühr et al., 2018; Willers et al., 2018) is located in structures responsible for neurocognition, PBT may even lead to an increase of such side effects (Jhaveri et al., 2018). In the context of this study, premorbid intellectual abilities could not be estimated as neurocognitive test results were not available for the period prior to tumour diagnosis (as usual for clinical studies). Since information about the years of education was missing, an education correction could not be applied (Nasreddine et al., 2005; Thomann et al., 2018). Therefore, this study focused on changes in MoCA scores and only patients with available baseline values were included. Since deterioration of the MoCA score was observed in only a few patients, the development of NTCP models including dosimetric parameters as predictors based on this data was not feasible.

5.4 Summary

In this study, self-reported and objectively measured neurocognition and most other QoL domains remained largely stable over time during recurrence-free follow-up in brain tumour patients treated with PBT. Slight deterioration of the MoCA score was associated with an increased relative volume of the irradiated anterior cerebellum exposed to doses of 30 to 40 Gy(RBE). Larger studies, such as the ongoing comparative clinical trials (NCT03180502, 2017) and (NCT02824731, 2016) may allow for validation of these findings, development of NTCP models and a direct comparison to neurocognitive changes of patients treated with either cranial PBT or XRT.

Future studies on radiation-induced neurocognitive dysfunction may profit from more comprehensive standardised test batteries. The detailed follow-up may also comprise specific questionnaires to address the subjective perception of neurocognitive function, e.g. the FACT-Cog questionnaire. This extended testing procedure could be performed less frequently, e.g. yearly, for a follow-up of several years. Furthermore, additional patient-specific information, e.g. years of education, should be obtained.

6 Treatment plan and NTCP comparison for patients with intracranial tumours

6.1 Motivation

The treatment decision for PBT is often justified by the superior dose distribution of PBT compared to XRT. Patient-individual *in silico* comparison of treatment plans of both techniques may provide the rationale for treatment decisions. The model-based approach for patient selection, however, goes beyond a purely physical treatment plan comparison and converts the dosimetric differences into side effect probabilities.

This chapter retrospectively investigates and quantifies the dosimetric benefit of PBT in terms of dose distribution compared to XRT for several brain structures on a patient cohort treated with PBT in the first section. The potential benefit in terms of reduced probabilities of radiation-induced side effects is estimated based on a variety of NTCP models in the second section.

6.2 Treatment plan comparison of cranial proton and photon radiotherapy

6.2.1 Patient cohort and experimental design

Patient data

The patient cohort included two subgroups of patients with intracranial tumours characterised in detail in chapter 3. Subgroup 1 comprised 117 patients treated between 12/2014 and 11/2016 at UPTD. Subgroup 2 enclosed 67 patients treated at WPE between 08/2013 and 08/2016. Patient data from MGH were not considered for dose comparison as no general dose constraints but patient-specific treatment constraints were applied for PBT planning. Inclusion criteria were diagnosis of LGG, HGG, meningioma, pituitary adenoma or craniopharyngioma, age $\geq$ 18 years and ECOG performance status $\leq$2, normo-fractionated PBT as well as written informed consent. Exclusion criteria were inability to MRI planning, lack of patient compliance and lack of or limited possibilities for reproducible positioning. A further exclusion criterion was a treatment scheme with sequential boost (45 patients). As this dose-comparison study investigated the dose differences in normal tissues, equivalent target coverage for the PBT and the XRT treatment plan had to be ensured. Therefore, further strict CTV coverage criteria were applied for both treatment plans (described in more detail below). After this selection, 92 patients remained in the cohort which is further characterised in table 6.1. This retrospective study was approved by the local Ethics Committee (EK566122019).

Table 6.1: Patient characteristics of the dose-comparison study.

Characteristic	**Median**	**Range**
Age at PBT in years	45.6	18.1 – 84.7
Tumour volume (CTV) in cm^3	93.6	1.2 – 396.1
Prescribed dose in Gy(RBE)	54.0	30.0 – 60.0
	N	**(%)**
Gender [female/male]	45/47	(48.9/51.1)
Surgery [yes/no]	75/17	(81.5/18.5)
Chemotherapy [yes/no]	25/67	(27.2/72.8)
PBT technique [DS/PBS]	45/47	(48.9/51.1)
Re-irradiation [yes/no/missing]	7/82/3	(7.6/89.1/3.3)
Tumour location		
Temporal	25	(27.2)
Frontal	16	(17.4)
Parietal	5	(5.4)
Occipital	4	(4.3)
$>$1 lobe	31	(33.7)
Other	11	(12.0)
Tumour histology		
High grade glioma	31	(33.7)
Low grade glioma	14	(15.2)
Meningioma	22	(23.9)
Other	25	(27.2)

Abbreviations: CTV, clinical target volume; N, number of patients; DS double scattering; PBS, pencil beam scanning.

Delineation of target volumes and organs at risk

For delineation, a CT scan with a slice thickness of 2 mm (subgroup 1) or 1 mm (subgroup 2) was acquired for each patient and rigidly coregistered with pre-therapeutic MRI scans (T1-weighted, T2-weighted or fluid-attenuated inversion recovery (FLAIR) with and without contrast agent). Target delineation was performed in clinical routine as described in section 2.2.4. The GTV comprised the macroscopic tumour or the resection cavity with residual tumour volume, if present. The CTV included microscopic disease and oedema and was created based on the GTV or resection cavity taking into account information from MRI and the original tumour histology as well as anatomical boundaries.

Additional OARs that were not contoured and considered during treatment planning for clinical application were retrospectively delineated on the original treatment planning CT and the coregistered MRI scans. These additional structures included hippocampi, cerebellum (posterior, anterior and entire organ), frontal and temporal lobes using organ-specific contouring guidelines (Schmahmann et al., 1999; Chera et al., 2009; Sun et al., 2014; Scoccianti et al., 2015; Eekers et al., 2018), see table 6.2. The brain stem consisted of the medulla oblongata, pons and midbrain. The brain included infra- and supratentorial structures excluding the brain stem. The remaining brain (brain-CTV), created by subtracting the CTV from the brain structure, was considered for evaluation. In paired organs, the organ with a higher mean dose was defined as ipsilateral and

Table 6.2: Considered organs at risk and dose-volume parameters for comparison of treatment modalities.

Structure	Dose-volume parameter
Brain stem	$D2\%$
Brain-CTV	$D2\%$, $D50\%$, D_{mean}, $V10Gy(RBE)$, $V20Gy(RBE)$, ..., $V60Gy(RBE)$ in % of entire volume
Cerebellum	$D2\%$, $D50\%$, D_{mean}, $V10Gy(RBE)$, $V20Gy(RBE)$, ..., $V60Gy(RBE)$ in % of entire volume
Cerebellum anterior	$D2\%$, $D50\%$, D_{mean}, $V10Gy(RBE)$, $V20Gy(RBE)$, ..., $V60Gy(RBE)$ in % of entire volume
Cerebellum posterior	$D2\%$, $D50\%$, D_{mean}, $V10Gy(RBE)$, $V20Gy(RBE)$, ..., $V60Gy(RBE)$ in % of entire volume
Chiasm	$D2\%$
Cochlea*	$D2\%$, D_{mean}
Frontal lobe*	$D2\%$, $D50\%$, D_{mean}, $V10Gy(RBE)$, $V20Gy(RBE)$, ..., $V60Gy(RBE)$ in % of entire volume
Hippocampus*	$D2\%$, $D50\%$, D_{mean}
Lacrimal gland*	$D2\%$, D_{mean}
Lens*	$D2\%$, D_{mean}
Optic nerve*	$D2\%$
Pituitary	$D2\%$, D_{mean}
Skin	$D2\%$, $D5\%$, $D15\%$..., $D55\%$, $V10Gy(RBE)$, $V20Gy(RBE)$, ..., $V60Gy(RBE)$ in cm^3
Temporal lobe*	$D2\%$, $D50\%$, D_{mean}, $V10Gy(RBE)$, $V20Gy(RBE)$, ..., $V60Gy(RBE)$ in % of entire volume

*ipsi- and contralateral organs separately

the remaining organ as contralateral. The contra- and ipsilateral hippocampus as well as the bilateral hippocampi were considered. Structures located directly within the macroscopic tumour volume were not contoured. Delineation was performed using RayStation®.

Treatment planning

The clinically applied PBT treatment plan was evaluated. The XRT plan was retrospectively created using a state-of-the-art VMAT technique.

Proton treatment planning Both centres were equipped with an IBA (Ion Beam Applications S.A., Louvain-La-Neuve, Belgium) system utilised with an isochronous cyclotron and an isocentric gantry. Some patients of subgroup 2 were treated at a fixed beamline.

For subgroup 1, treatment plans were calculated using a DS technique and the treatment planning system XiO®. The target volume was the CTV with a relative in-beam margin of 3.5 % of the distal or proximal CTV depth and an additional absolute margin of 3 mm. The treatment planning was based on the beam data of an IBA universal nozzle. The highest available proton energy was 230 MeV. A range shifter was inserted to enable the dose application close to the patient surface. The air gap between range shifter and the patient surface was set to 5 cm. Field-specific brass apertures and milled lucite range compensators were created to conform the dose to the target. A $3 \times 3 \times 3\,mm^3$ dose calculation grid and a pencil-beam dose algorithm was used.

For subgroup 2, a PBS technique was applied. Treatment plans were generated using the treatment planning systems XiO® (2013 – 2014) and RayStation® (2015 – 2016). Treatment was planned on a PTV constructed by adding a 5 mm margin to the CTV. For both treatment planning systems XiO® and RayStation®, treatment planning was based on beam data of the

IBA dedicated nozzle. For patients treated at a fixed beamline, an IBA universal nozzle was used. The sigma of the Gaussian spot size varied between 2.7 and 3.1 mm in air at isocentre for the highest proton energy depending on the treatment room. A regular spot grid with a variable spot spacing between 0.8 and 1.0 cm was used. The energy layer distance was automatically selected by the treatment planning system. The minimal required number of monitor units per spot was set at 0.04. For the treatment of superficial tumours, a range shifter was inserted. The fixed air gap between the range shifter and the isocentre was 43 cm for the dedicated nozzle. For the universal nozzle, a fixed air gap of 5 cm between range shifter and the patient surface was applied if there was no risk of collision. Dose calculation was performed based on a pencil beam algorithm with a dose calculation grid of $1 \times 1 \times 1\,\mathrm{mm}^3$.

One to four beams in a patient individual beam setting (coplanar or non-coplanar) could be applied according to clinical practice in the respective centres. The RBE was considered to be constant at 1.1. To account for this dose weighting, the unit Gy(RBE) is used for both treatment modalities throughout this dose-comparison study. The dose-volume constraints differed slightly between both subgroups, see appendix B tables B.4 and B.5.

Photon treatment planning Treatment planning of VMAT was performed using the planning system RayStation® with a collapsed-cone algorithm. Treatment plans were conformed to a PTV created by adding a uniform margin of 5 mm to the CTV. Two gantry rotations were applied (clockwise and counter-clockwise). Start and stop angles were slightly adapted to avoid unnecessary exposure of normal tissue for small unilateral located volumes. Coplanar arcs were used in the majority of patients. Angles of the multi-leaf collimator were 45° and 315°, respectively. A photon energy of 6 MV was used. Treatment planning was based on beam data of a Varian linear accelerator 2100 (Varian Medical Systems, Palo Alto, CA, USA) with a dynamic multileaf collimator (dMLC) with 60 leaf pairs at a width of 5 mm and 10 mm for the central 40 pairs and the remaining, respectively. The maximum field size was $40 \times 40\,\mathrm{cm}^2$. The arc gantry spacing between the control points of the sliding window segments was set to 4°, the maximum gantry angle speed was $5.54^\circ\mathrm{s}^{-1}$. The dose calculation grid was set to $1 \times 1 \times 1\,\mathrm{mm}^3$. All evaluations were performed for the calculated nominal dose distributions.

Treatment planning aimed for clinically acceptable treatment plans in terms of target coverage and dose constraints for OARs. To assure comparability, the same OARs and dose constraints as in PBT treatment planning were applied, see appendix B tables B.4 and B.5. At least 95 % of the target (PTVs) should receive 95 % of the prescribed dose ($V95\% > 95\,\%$). Volumes receiving more than 107 % of the prescribed dose should be avoided or minimised ($D2\% < 107\,\%$). The brain stem was considered as an avoidance structure with higher priority than the target coverage. Thus, the planning goals for the target volumes could be compromised for the sake of brain stem sparing.

Treatment plan evaluation

The retrospectively created XRT treatment plans were qualitatively evaluated by one experienced radiation oncologist. This was done additionally to the above-mentioned quantitative parameters (e.g. $V95\%$, $D2\%$), as these simple metrics cannot reflect the three-dimensional dose distribution exactly. Hence, treatment plans that did not pass the evaluation for clinical acceptability were re-planned or dismissed although they met the nominal criteria. All PBT plans had been applied clinically and thus were approved by experienced radiation oncologists during clinical routine.

All data were derived from patients that had already been treated with PBT. To allow for a comparison between related PBT and XRT treatment plans, similar target coverage had to be ensured. For this sake, further strict selection criteria for the CTV were applied to each plan: $D98\% \geq 95\,\%$ and $D95\% \geq 98\,\%$ of the prescribed dose. These target coverage criteria were not met in 9, 24 and 14 patients for the XRT, PBT or both treatment plans, respectively. Finally, 92 patients were included in the dose-comparison study for which both treatment plans met these strict requirements. One reason why the XRT plans did not meet the requirements could be that the patients were all selected and treated with PBT for various reasons. For example, because the target coverage may not be achieved with XRT. Patients with a tumour located very close to the brain stem were especially selected for PBT because the steep dose gradient was expected to provide better target volume coverage. For some PBT plans, the planning objectives were met, but not the stricter selection criteria for this plan comparison study.

Evaluation parameters All dose distributions were exported using RayStation® scripts. For comparison of the target coverage between XRT and PBT treatment plans, CTV coverage was evaluated due to the fact that DS PBT planning was performed based on CTVs rather than PTVs. The following metrics of the CTV were assessed: $D98\%$, $D2\%$, $V95\%$, $V105\%$, $V107\%$, the conformity index CI for 98 % of the prescribed dose, see equation 2.4 and the homogeneity index HI, see (2.5).

For the comparison of dose to OARs, different DVH parameters were selected. An overview is given in table 6.2. Absolute volume parameters (in cm^3) were considered for the non-circumscribed organ skin. For circumscribed OARs, relative volume parameters (in per cent of the entire volume of the structure) were assessed. For the brain stem and optic pathways, only the near-maximum dose was considered ($D2\%$). For OARs with volumes $< 10\,cm^3$, the near-maximum dose $D2\%$, D_{mean} and D_{median} ($D50\%$) were considered. As the median and mean dose are very similar for very small structures (lens, cochlea, lacrimal gland, pituitary gland), only $D2\%$ and D_{mean} were assessed.

Dosimetric analysis

Absolute differences in dosimetric parameters ΔD_{abs}, ΔV_{abs} between both treatment modalities XRT and PBT were assessed by

$$\Delta D_{\text{abs}} = D_{\text{XRT}} - D_{\text{PBT}} \tag{6.1}$$

$$\Delta V_{\text{abs}} = V_{\text{XRT}} - V_{\text{PBT}}, \tag{6.2}$$

where D_{XRT} and D_{PBT} represent dose parameters and V_{XRT} and V_{PBT} represent volume parameters of the XRT and PBT treatment plan, respectively. Relative differences in dosimetric parameters ΔD_{rel}, ΔV_{rel} between both treatment modalities XRT and PBT were expressed by

$$\Delta D_{\text{rel}} = \frac{D_{\text{XRT}} - D_{\text{PBT}}}{D_{\text{XRT}}} = 1 - \frac{D_{\text{PBT}}}{D_{\text{XRT}}} \tag{6.3}$$

$$\Delta V_{\text{rel}} = \frac{V_{\text{XRT}} - V_{\text{PBT}}}{V_{\text{XRT}}} = 1 - \frac{V_{\text{PBT}}}{V_{\text{XRT}}}. \tag{6.4}$$

The patient-individual differences between the XRT and PBT plan were assessed by a Wilcoxon signed-rank test. Moreover, to evaluate the extent of the dose sparing that can be achieved using PBT instead of XRT, a threshold of 10 % was applied to individual dosimetric parameters that were suggested to be clinically relevant by Scoccianti et al. (2015). These comprise mean doses of the hippocampi and cochlea as well as maximum doses of the brain stem, optical pathways, lacrimal gland, lens and pituitary. Furthermore, a dosimetric comparison for patient groups classified according to the tumour location (frontal, temporal, parietal and multiple affected lobes) was conducted.

6.2.2 Results

Target coverage

The CTV coverage was similar for both treatment techniques, see table 6.3. The high-dose parameter $V107\%$ was below 0.15 % for all treatment plans and thus, not further evaluated. No statistically significant difference between XRT and PBT could be observed for $D98\%$, $D2\%$ and $V105\%$. Even though the difference between both techniques for $V95\%$ was statistically significant, the absolute difference was very small (< 1 percentage point). Conformity was better for the PBT plans and homogeneity was marginally better for XRT plans.

Dose comparison for organs at risk

All investigated dose-volume parameters for all considered OARs are given in appendix D table D.1 for both XRT and PBT treatment plans. For almost all investigated structures and dosimetric parameters, PBT reduced the dose significantly compared to XRT. No significant difference

Table 6.3: Comparison of parameters characterising the target coverage. Volume parameters *Vx*% are given in per cent of the volume of the CTV, dose parameters *Dx*% are given in per cent of the prescribed dose. Mean values and standard deviations are presented. P-values were derived from Wilcoxon signed-rank tests.

Parameter	PBT			XRT			XRT – PBT			p-value
*D*98% in %	98.6	±	0.8	98.7	±	0.7	0.16	±	0.87	0.12
*D*2% in %	102.4	±	0.9	102.2	±	0.9	-0.17	±	1.29	0.13
*V*95% in %	99.9	±	0.3	99.9	±	0.3	0.04	±	0.26	0.006
*V*105% in %	0.1	±	0.3	0.1	±	0.3	0.02	±	0.48	0.89
*V*107% in %	0.0	±	0.0	0.0	±	0.0	0.00	±	0.02	0.062
HI in %	3.8	±	1.2	3.5	±	1.2	-0.33	±	1.58	0.028
*CI*98% in %	49.9	±	14.7	43.3	±	15.1	-6.63	±	6.61	$<$0.001

Abbreviations: HI, homogeneity index; *CI*98%, conformity index for 98 % isodose.

could be observed for the parameter $V60Gy(RBE)$ describing the very high dose region. For both treatment techniques, the relative values of $V60Gy(RBE)$ in the temporal and frontal lobes, cerebellar structures and brain-CTV were very low (mean values below 3.5 %). Similarly, the absolute values of $V60Gy(RBE)$ of the skin were below 0.5 cm^3. Dose reduction using PBT instead of XRT was mainly observed in the low dose region described by $V10Gy(RBE)$ or $V20Gy(RBE)$.

Figure 6.1 exemplarily shows selected volume parameters of brain-CTV and skin for both treatment modalities. For brain-CTV, almost all dosimetric parameters were significantly lower for PBT compared to XRT. For the skin, however, only low to intermediate dose areas were significantly reduced for PBT compared to XRT. The volume of the skin receiving high doses described by $V40Gy(RBE)$ to $V50Gy(RBE)$ was smaller for XRT compared to PBT, however not significantly. This effect can be illustrated by the build-up region for photon beams that is not present for proton beams.

In paired organs, PBT allowed greater sparing of the contralateral than the ipsilateral organ compared to XRT. This is shown exemplarily for the hippocampus in figure 6.2. While PBT could reduce the near-maximum dose parameter *D*2% of the ipsilateral hippocampus in 80.8 % of the patients, a reduction of this parameter could be achieved in 91.8 % of the patients for the contralateral hippocampus. For even more patients, the mean hippocampal dose could be reduced by using PBT compared to XRT. This was the case in 89.7 % of the patients for the ipsilateral and in 96.5 % of the patients for the contralateral hippocampus. In a large number of patients, PBT deposited almost no dose in the contralateral hippocampus, while XRT delivered mean doses up to 35 Gy(RBE).

Patient-individual dosimetric differences between both treatment modalities were also calculated for selected DVH parameters for intracranial OARs suggested by Scoccianti et al. (2015). These relative differences are presented in appendix D figure D.1 for each patient. Appendix D figure D.2 shows the parameters exceeding a threshold of 10 %. It demonstrates that most parameters of all patients can be reduced by more than 10 % when using PBT instead of XRT. However, the relative difference does not indicate the extent of the reduction and whether this

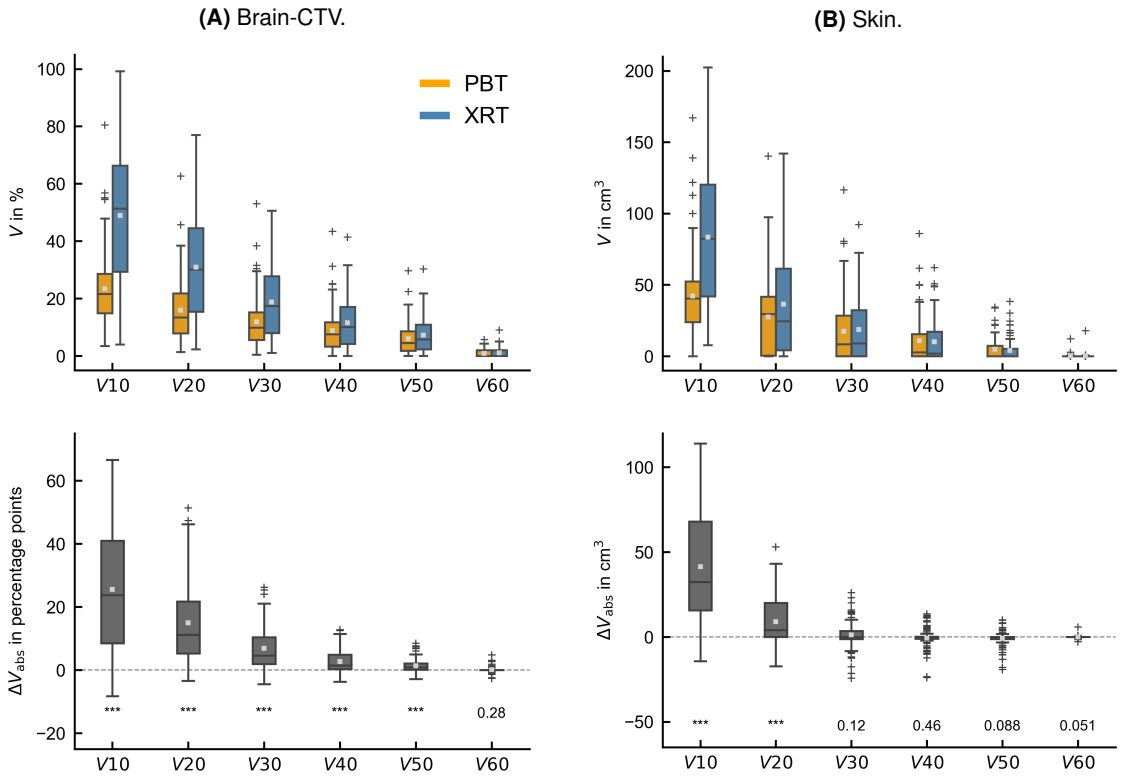


Figure 6.1: Volume parameters *VxGy*(*RBE*) abbreviated as *Vx* of brain-CTV **(A)** and skin **(B)** for XRT and PBT treatment plans. The upper panel shows absolute values, the bottom panel shows the difference between XRT and PBT values. P-values of the paired Wilcoxon test are given; *** represents $p < 0.001$.

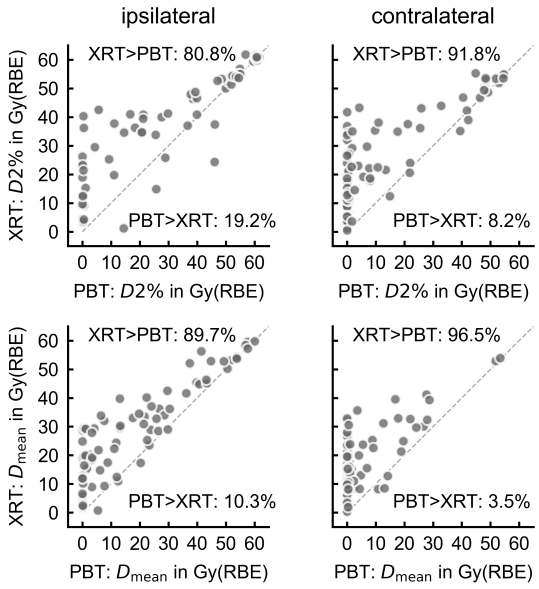


Figure 6.2: Dosimetric parameters of the ipsilateral (left panel) and contralateral hippocampus (right panel). The dashed line represents equal values for both treatment modalities. The percentage of patients above or below this line is indicated.

reduction is also clinically relevant. This will be explained in the following using two example patients. For these two exemplary patients, the dose distributions for both treatment modalities are displayed on CT images in figure 6.3. The characteristics of these patients are given in table 6.4.

Figure 6.4 shows absolute differences and classified relative differences between XRT and PBT as well as the absolute dose values for the above-mentioned parameters for both example patients. For patient 1, the relative reduction using PBT instead of XRT is larger than 10 % for all selected parameters. PBT is able to reduce a number of dosimetric parameters to zero, while the corresponding doses for XRT are between 2 and 20 Gy(RBE), i.e. also for XRT the general level of dose exposure is comparatively low. For patient 2, only a few parameters could be reduced

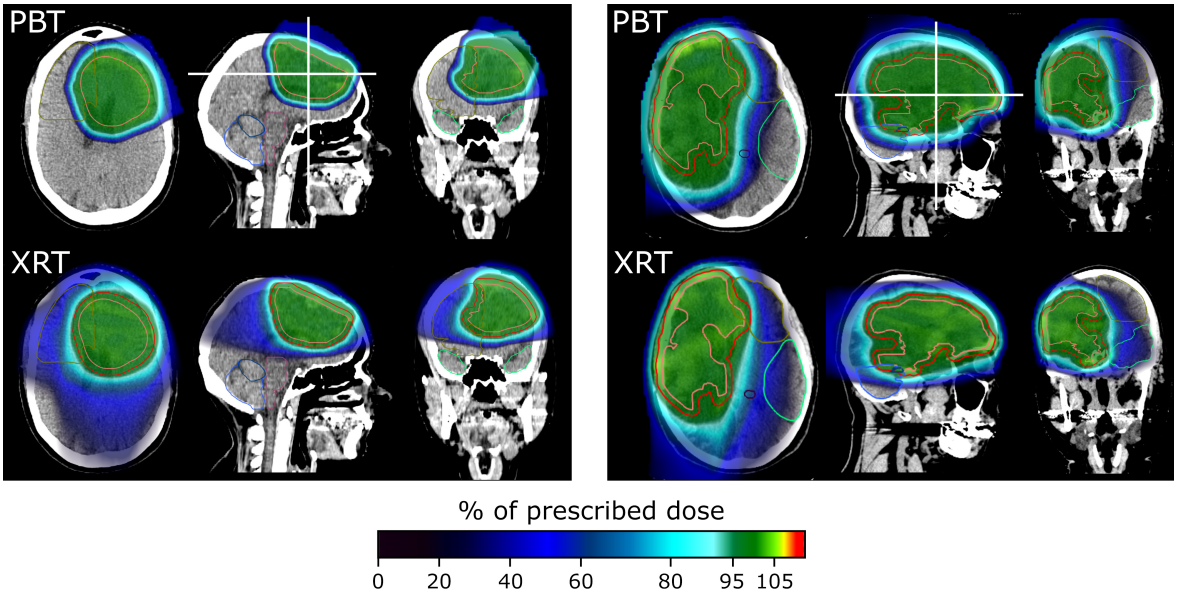


Figure 6.3: Dose distributions for both treatment modalities (PBT and XRT) of two example patients. Doses are given in per cent of the prescribed dose (60 Gy(RBE) for patient 1 and 54 Gy(RBE) for patient 2). The white lines in the sagittal view represent the sectional planes of the coronal and axial views.

Table 6.4: Characteristics of the example patients.

Characteristic	**Patient 1**	**Patient 2**
Gender	Female	Female
Tumour histology	Oligoastrocytoma grade III	Astrocytoma grade II
Tumour location	Left-sided	Right-sided
	Frontal lobe	$>$ 1 lobe involved
Surgery	Yes	Yes
CTx	Yes	No
PBT technique	DS	PBS
Age in years	48	61
Tumour volume (CTV) in cm^3	117	240
Prescribed dose in Gy(RBE)	60	54

Abbreviations: CTV, clinical target volume; CTx, chemotherapy; DS, double scattering; PBS, pencil beam scanning.

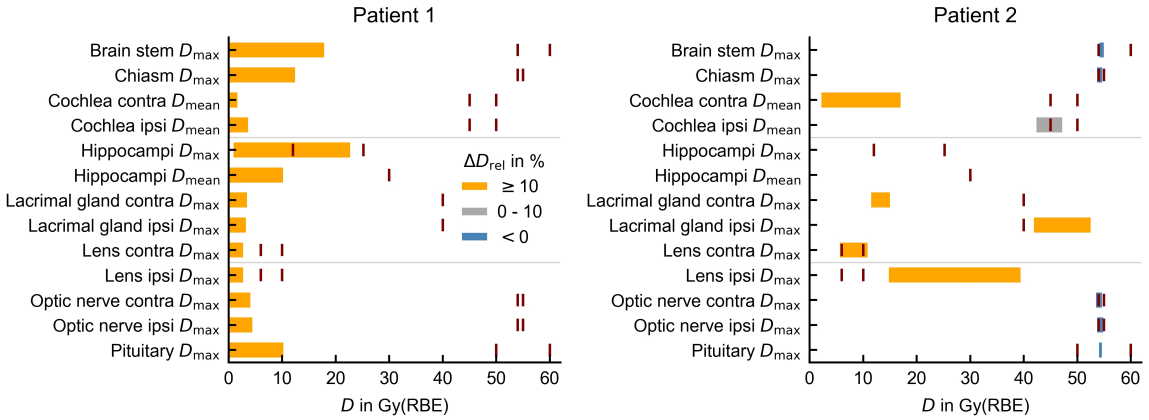


Figure 6.4: Relative and absolute difference of selected DVH parameters between PBT and XRT for two example patients. The absolute PBT and XRT values of the respective dosimetric parameters represent the left and right limits of the bars. The shading indicates whether the XRT plan revealed lower doses than the PBT plan (relative difference ΔD_{rel} below 0 %, blue), or the PBT plan revealed lower doses than the XRT plan (relative difference ΔD_{rel} above 0 %, grey). If the relative improvement using PBT was greater than 10 %, the bar is shaded in orange. Parameter selection and dose constraints (given as red lines, multiple constraints for some parameters) as suggested by Scoccianti et al. (2015). ipsi, ipsilateral; contra, contralateral.

by more than 10 % using PBT instead of XRT. However, the dose to the ipsilateral lens could be reduced substantially. Also, dosimetric parameters of the lacrimal glands as well as contralateral lens and cochlea could be reduced by more than 10 % using PBT. For the mean dose of the ipsilateral cochlea, the reduction remained below 10 % as the dose exposure was relatively high for both treatment modalities ($>$ 40 Gy(RBE)). Though being rather moderate, this reduction may be relevant as cochlea dose is very close to its constraints. The maximum doses in the brain stem, optic pathways and pituitary were minimally reduced with the XRT plan. The dose to the bilateral hippocampi was not shown, as the ipsilateral hippocampus was completely covered by the tumour.

An analysis of differences in dosimetric parameters (D_{mean} and $D2\%$) between the two treatment modalities for different tumour localisations (frontal, temporal, parietal and multiple affected lobes) is presented in appendix D table D.2. The highest relative dose reduction could be observed for tumours located in the parietal lobe. The median relative reduction for the majority of OARs was above 50 %, except for ipsilateral hippocampus and temporal lobe. The greatest median absolute reduction was observed for $D2\%$ of the contralateral frontal temporal lobe (30.2 Gy(RBE)). However, only five patients had a tumour location restricted to the parietal lobe. For patients with a tumour located in the frontal lobe, a median relative reduction of 100 % for cerebellar doses could be observed using PBT. The greatest median absolute reduction was seen for $D2\%$ of the brain stem (15.7 Gy(RBE)). The dose reduction using PBT was slightly less pronounced in temporal tumours. The contralateral lacrimal gland, lens and cochlea could be spared by 100 % in the median. The greatest median absolute reduction was observed for $D2\%$

of the contralateral hippocampus (15.8 Gy(RBE)) using PBT instead of XRT. For patients with a tumour located in more than one lobe, a large relative sparing ($> 90\,\%$) could be observed especially in contralateral structures (cochlea, lacrimal gland, lens, temporal lobe, hippocampus). The greatest median absolute reduction was observed for *D*2% of the contralateral hippocampus (14.1 Gy(RBE)).

6.2.3 Discussion

PBT was able to reduce the dose in almost all investigated OARs and dosimetric parameters substantially, especially in the low and intermediate dose range and for contralateral organs. The only exception was the skin, where the dose in the intermediate to high range was even slightly increased for PBT. In contrast to photon beams, proton beams show no build-up effect, leading to higher skin doses.

The presented results are in line with other dose-comparison studies. However, most of the publications comprised only a small number of patients and did not investigate the skin as an OAR. Baumert et al. (2001) compared stereotactic XRT with a micro-multileaf collimator and PBS PBT in seven patients with differently shaped targets (concave, ellipsoid isolated, superficial and close to an OAR as well as irregular complex). They found no benefit for PBT compared to stereotactic XRT in simple shaped or superficial tumours. For complex targets or targets located near critical structures, PBT seemed to be a potentially better choice than XRT. Bolsi et al. (2003) compared three different XRT techniques (3D-CRT, stereotactic arc therapy and IMRT) and two different PBT techniques (PBS and DS) in a cohort of twelve patients with benign brain tumours. Both PBT techniques had superior dosimetric properties compared to all XRT techniques concerning target dose uniformity and conformity and in terms of sparing OARs. The authors even proposed to consider PBT as the treatment of choice for patients with a long life expectancy.

Dennis et al. (2013) conducted a dose-comparison study between IMRT and DS PBT in eleven LGG patients. They estimated a halving of *gEUD* in healthy tissues surrounding the tumour using PBT instead of IMRT. Moreover, they stated that the risk of secondary tumour induction after IMRT would be twice as high as with PBT. As in this thesis, the authors concluded that PBT was dosimetrically superior to XRT, especially for critical structures that were located on the contralateral side compared to the tumour.

Adeberg et al. (2016) compared intensity-modulated PBT, VMAT and 3D-CRT in a small cohort of twelve patients with HGG. They also found a considerable dose reduction for PBT, especially in contralateral organs of neurocognitive function and in areas of neurogenesis. Harrabi et al. (2016) investigated the dose distribution of intensity-modulated PBT and 3D-CRT plans in 74 HGG patients. They concluded similarly that PBT is dosimetrically superior to conventional XRT, especially for OARs that are considered critical for neurological function and neurocognition or that play an important role in the preservation of QoL.

Adeberg et al. (2018) performed a treatment plan comparison between PBT, 3D-CRT and VMAT to identify patient subgroups based on tumour localisation that may benefit most from PBT. They investigated relative dose differences in 50 patients with five different tumour localisations (frontal, suprasellar, temporal, parietal and posterior fossa). Patients with parietal tumours exhibited the greatest relative dose reductions in most OARs, especially in the optical pathways and contralateral OARs. For tumours located in the temporal lobe, the reductions were prominent in the mean dose of the infra- and supratentorial brain as well as the whole brain. For frontal lobe tumours, mainly the mean dose of the infratentorial brain and contralateral hippocampus could be reduced. These results described by Adeberg et al. (2018) are consistent with the results of the dose-comparison study presented in this thesis. However, the dosimetric benefit refers to the relative reduction and absolute dose differences were not analysed. In this context, some relative dose reductions could be overrated. For example, a 100 per cent relative reduction can easily be achieved if no dose at all is deposited in an OAR by PBT and only very low doses are delivered by XRT. Adeberg et al. (2018) concluded that their results could improve the selection of patients who should receive PBT, as patients with rather small tumours near critical structures could benefit most from PBT with regard to dosimetric benefits. Nevertheless, the classification into one of the five localisation groups does not necessarily allow for conclusions on the benefit of an individual patient. As an example, the authors suggest that the dosimetric benefit of PBT for contralateral OARs of a midline and a well-lateralised frontal lobe lesion may differ considerably, although both tumours would be assigned to the same localisation group.

This dose-comparison study showed that the highest relative dose reduction can be achieved in organs that are at a greater distance from the target (contralateral). Since PBT deposits almost no dose in distant organs, the relative reductions compared to XRT may be substantial, although the absolute XRT doses also may be relatively low (see example patient 1). Thus, a state-of-the-art XRT technique may be a valuable option for these patients, too. In some patients the tumour is very space-occupying, sometimes it covers large parts of one hemisphere (the maximum tumour volume in this study was almost 400 cm^3). In patients with large tumours, even the contralateral organs may be located in the near to middle distance and can hardly be spared with either technique (see example patient 2). This suggests, that patients with a very large tumour may benefit less from PBT compared to patients with a rather small tumour and may be treated with a state-of-the-art XRT technique. However, no correlation (with Spearman $|\rho| > 0.5$) could be observed between increasing tumour volume (CTV) and decreasing relative and absolute dosimetric differences for all investigated OARs. This indicates that, in general, the extent of a dosimetric benefit cannot be determined solely on the basis of tumour location or size, but that a patient-specific assessment is required.

In general, clinical decisions to treat certain brain tumour patients with PBT are based on a comparison of dose distributions rather than a detailed comparison of NTCP values. So far, the availability of validated NTCP models predicting side effects following cranial RT is limited. Thus, figures like figure 6.4 could guide physicians in the treatment decision, as relevant dosimetric pa-

rameters are displayed at a glance. For use in daily routine, however, the obstacle remains that clinically relevant OARs and dosimetric parameters must be defined. Furthermore, appropriate threshold values must be specified above which PBT is considered to be indicated. As shown above, a large relative difference in dosimetric parameters between treatment options does not necessarily have to be clinically relevant, so that both absolute differences and absolute values should be considered. Due to this multitude of degrees of freedom and decisions, the establishment of consistent guidelines (at least within one clinic) requires a great effort by clinicians in close cooperation with other professional groups. The most demanding challenge is the definition of the term *clinical relevance*. It is not defined so far since no results of prospective clinical studies comparing XRT and PBT for brain tumour patients are available yet (Adeberg et al., 2017). Furthermore, for each individual patient, sparing of OARs may be prioritised differently depending on patient-specific factors, e.g. comorbidities, age, life expectancy or patient preference. Hence, it is even more difficult to specify consistent selection criteria in clinical practice.

To reduce the number of degrees of freedom and to assess the quality of treatment plans for decision making, different single metrics have been developed including assessment of the target coverage as well as sparing of OARs. Target coverage metrics are for example the homogeneity or several conformity indices (Paddick, 2000; Feuvret et al., 2006; Yoon et al., 2007). To evaluate OAR sparing, the healthy tissue overdosage factor *HTOF* (Feuvret et al., 2006), critical organ-scoring index *COSI* (Menhel et al., 2006) or the healthy tissue conformity index *HTCI* (Feuvret et al., 2006) have been suggested. Most metrics have in common that they usually evaluate only one aspect and cannot assess the quality of the overall treatment plan. But there are also metrics that attempt to evaluate target coverage and OAR sparing simultaneously to provide overall quality scores (Miften et al., 2004; Akpati et al., 2008). A decision-support system to choose between different tentative treatment plans has been proposed by Alfonso et al. (2015) based on a dose distribution index *DDI*. It incorporates different metrics regarding the target as well as sparing of OARs and the remaining volume at risk. However, these more complex metrics are based only on physical parameters and are rarely implemented in clinical practice.

These difficulties, especially the clinical assessment of dose reduction, can be addressed if the dose parameters are directly converted into NTCP estimates. Whether the shown dosimetric superiority of PBT over XRT translates into a clinical benefit in terms of NTCP is assessed in the following section.

Limitations of the dose-comparison study

This treatment plan comparison was retrospectively conducted. Several OARs were retrospectively contoured and were therefore not included in the active planning process. As there are (still) no dose limits for many of these structures, they are not yet considered in clinical routine. The dose to OARs mainly depends on the applied treatment planning technique, beam arrange-

ment, treatment goals and objectives during the treatment planning optimisation process, which narrows the generalisability of all *in silico* dose-comparison studies.

The study cohort could be subject to a selection bias because the patients have already been treated with PBT and were selected for this treatment for certain reasons, e.g. sufficient target coverage may be not achieved using the available XRT techniques. However, this issue was reduced in the dose-comparison study by applying additional selection criteria to ensure comparable target coverage with both modalities. Although the patient cohort was rather heterogeneous, it reflects the wide range of possible clinical cases.

This dose-comparison study examined only the VMAT technique for XRT. Although 3D-CRT is also still common in clinical practice, this technique was not considered here as the study aimed at comparing PBT with a state-of-the-art XRT technique. Both PBT techniques, DS and PBS, were investigated, as patients included in this study were treated with DS at that time at the UPTD. Since the introduction of PBS at this centre, all patients with intracranial tumours have been treated with PBS. To investigate the impact of the PBT technique, the dosimetric differences between XRT and PBT were compared for the subgroups of patients treated with either PBT technique. For the skin, no significant differences between both patient groups could be observed for the absolute and relative dosimetric difference of a majority of parameters. For other parameters, the absolute and relative difference between XRT and PBT was on average higher in the group treated with DS compared to patients treated with PBS (e.g. $V10Gy(RBE)$). However, a comprehensive comparison between both PBT techniques, PBS and DS, requires an *in silico* treatment planning for each individual patient, as tumour-related parameters such as location, tumour size, prescribed dose and target volume concepts differed between both subgroups. This was out of the scope of this study comparing XRT and PBT in general.

Dose uncertainties in the dose calculation of RT treatment plans may have occurred due to the used dose calculation algorithms. The extent of these uncertainties mainly depends on the tissue heterogeneities in the beam path. They increase in areas with high tissue heterogeneity, for example at interfaces of air-filled nasal cavities, bone (skull) and soft tissues. Therefore, these dose calculation uncertainties are more likely in patients with a tumour in the skull base than in patients who do not require irradiation of the heterogeneous area (cranial lesions). Moreover, the dose uncertainties depend on the applied dose calculation algorithm. In this study, the clinically applied PBT treatment plans were calculated using a pencil beam algorithm. XRT plans were retrospectively calculated based on a collapsed cone algorithm. For XRT, this algorithm outperformed the pencil beam algorithm especially in the vicinity of interfaces and within low-density regions (Krieger and Sauer, 2005; Knöös et al., 2006). More accurate Monte Carlo dose calculation algorithms are becoming available in commercial treatment planning systems for both XRT and PBT. For PBT, a substantial impact of Monte Carlo algorithms compared to analytical algorithms is expected in complex anatomical regions (Paganetti, 2012). The evaluated dose differences in this study may be subject to uncertainties that arise from the different applied dose

calculation algorithms. However, the majority of patients in this study had a tumour located in the brain with less tissue heterogeneity.

6.3 Application of NTCP models

The results of the dose-comparison study in section 6.2 vividly confirmed that PBT plans can spare dose to healthy tissues compared to XRT, especially in the low to intermediate dose range and in contralateral OARs. An estimation of whether this dosimetric superiority translates into differences in treatment-related side effects can be quantified using NTCP models. For the following feasibility study, available NTCP models were selected that predict side effects associated with cranial irradiation. However, multivariable, externally validated models for modern treatment techniques are still rare for these side effects. It is essential to bear in mind that the following models are subject to several uncertainties outlined in the discussion. Consequently, the NTCP values for each patient and each side effect are rather rough estimates.

6.3.1 Patient cohort and experimental design

Selection of NTCP models

As explained in section 2.4, models used for patient selection need to be validated externally in terms of calibration and discrimination. Ideally, NTCP models should be derived from similar patient populations regarding tumour site, RT technique, and concomitant treatments, etc. However, for brain tumour patients, validated NTCP models for modern techniques are rare. This is due to the rather unspecific nature of common side effects (e.g. fatigue) and due to difficulties in their measurement such as for neurocognitive sequelae or radiation-induced brain necrosis. Nevertheless, different NTCP models were selected, that describe several radiation-induced side effects following treatments of tumours of the brain and skull base. The models were derived from normo-fractionated treatment data or use a correction for other fractionation schemes (e.g. EQD_2). Eleven different side effects were investigated: acute alopecia, acute erythema, blindness, brain and brain stem necrosis, cataract requiring intervention, delayed recall, endocrine dysfunction, hearing loss, ocular toxicity, temporal lobe injury and tinnitus (Burman et al., 1991; Bender, 2012; Gondi et al., 2012; Batth et al., 2013; De Marzi et al., 2015; Lee et al., 2015; Kong et al., 2016; Dutz et al., 2019). They comprise severe toxicities such as brain necrosis and rather less severe side effects such as alopecia, which can even though impact patients' quality of life. Both acute and late side effects were considered. Side effects that may occur following irradiation of one of a paired organ (e.g. cataract, hearing loss or tinnitus) were investigated for the contra- and ipsilateral OAR separately. The investigated NTCP models, parameters and coefficients are presented in table 6.5. They comprise different model types such as logistic models and LKB models using *gEUD* as a model predictor (see equation 2.12). The investigated side effects and selected models are presented in alphabetic order in the following.

Alopecia Acute alopecia is a common side effect following cranial RT that may affect patients' QoL. A logistic model for acute alopecia grade ≥ 2 following PBT including $D5\%$ of the skin as model predictor was applied as comprehensively described in section 4.2 (Dutz et al., 2019).

Blindness Irradiation of the optic nerves and chiasm may lead to visual loss as a severe late side effect of cranial RT. This side effect may occur in patients treated for tumours located in the vicinity of optic pathways (e.g. orbit, paranasal sinuses, skull base or frontal lobes). Loss of vision usually occurs 18 months after therapy with irradiation of more than 50 Gy (Danesh-Meyer, 2010; Thakkar et al., 2017). If irradiation is unavoidable, the risk of visual loss may be accepted by many patients to eliminate the tumour. However, since vision loss is difficult to treat, the risk of

Table 6.5: Considered NTCP models for potential side effects following cranial radiotherapy.

Endpoint	NTCP model	Model coefficients
Alopecia[1] **grade** ≥ 2 Acute Dutz et al. 2019	$NTCP = \left(1 + e^{-\beta_0 - \beta_1 D5\%}\right)^{-1}$	Skin $D5\%$ $\beta_0 = -1.33$, $\beta_1 = 0.08\,\mathrm{Gy(RBE)}^{-1}$
Blindness 5 years post-RT Burmann et al. 1991	$NTCP = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \exp\left(\frac{-u^2}{2}\right)\mathrm{d}u,\ t = \frac{gEUD - TD_{50}}{m \cdot TD_{50}}$	Chiasm and optic nerves $gEUD$ $TD_{50} = 65.0\,\mathrm{Gy}$, $m = 0.14$, $a = 4.0$
Brain necrosis 5 years post-RT Bender et al. 2012	$NTCP = \left(1 + \left(\frac{D_{50}}{EQD_2}\right)^{4\gamma}\right)^{-1}$	Brain-CTV and brain stem $D_{\max}$ (EQD_2) $D_{50} = 109.0\,\mathrm{Gy}$, $\gamma = 2.8$, $\alpha/\beta = 0.96\,\mathrm{Gy}$
Cataract requiring intervention 5 years post-RT Burmann et al. 1991	$NTCP = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \exp\left(\frac{-u^2}{2}\right)\mathrm{d}u,\ t = \frac{gEUD - TD_{50}}{m \cdot TD_{50}}$	Lenses $gEUD$ $TD_{50} = 18.0\,\mathrm{Gy}$, $m = 0.27$, $a = 3.33$
Delayed recall[2] 1.5 years post-RT Gondi et al. 2012	$NTCP = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \exp\left(\frac{-u^2}{2}\right)\mathrm{d}u,\ t = \frac{EQD_2 - EQD_2^{50}}{m \cdot EQD_2^{50}}$	Bilateral hippocampi $D40\%$ (EQD_2) $EQD_2^{50} = 14.88\,\mathrm{Gy}$, $m = 0.540$, $\alpha/\beta = 2\,\mathrm{Gy}$
Endocrine dysfunction[1] At least 0.5 – 2 years post-RT De Marzi et al. 2015	$NTCP = \left(1 + \left(\frac{TD_{50}}{gEUD}\right)^{4\gamma_{50}}\right)^{-1}$	Pituitary $gEUD$ $TD_{50} = 60.5\,\mathrm{Gy}$, $\gamma_{50} = 5.2$, $a = 6.4$
Erythema[1] **grade** ≥ 2 Acute Dutz et al. 2019	$NTCP = \left(1 + e^{-\beta_0 - \beta_1 V35Gy(RBE)}\right)^{-1}$	Skin $V35Gy(RBE)$, absolute volume $\beta_0 = -1.54$, $\beta_1 = 0.06\,\mathrm{cm}^{-3}$
Hearing loss[1] At least 0.5 – 2 years post-RT De Marzi et al. 2015	$NTCP = \left(1 + \left(\frac{TD_{50}}{gEUD}\right)^{4\gamma_{50}}\right)^{-1}$	Cochlea $gEUD$ $TD_{50} = 56.0\,\mathrm{Gy}$, $\gamma_{50} = 2.9$ $a = 1.2$
Ocular toxicity[3] **grade** ≥ 2 Acute Batth et al. 2013	$NTCP = \left(1 + e^{-\beta_0 - \beta_1 D_{\max}}\right)^{-1}$	Ipsilateral lacrimal gland $D_{\max}^4$ $\beta_0 = -5.174$, $\beta_1 = 0.205\,\mathrm{Gy}^{-1}$
Temporal lobe injury 5 years post-RT Kong et al. 2016	$NTCP = \left(1 + e^{-\beta_0 - \beta_1 D_{\max}}\right)^{-1}$	Temporal lobe $D_{\max}^4$ $\beta_0 = -18.61$, $\beta_1 = 0.227\,\mathrm{Gy}^{-1}$
Tinnitus[5] **grade** ≥ 2 1–2 years post-RT Lee et al. 2015	$NTCP = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \exp\left(\frac{-u^2}{2}\right)\mathrm{d}u,\ t = \frac{D_{\mathrm{mean}} - TD_{50}}{m \cdot TD_{50}}$	Cochlea D_{mean} $TD_{50} = 46.52\,\mathrm{Gy}$, $m = 0.35$

[1]CTCAE, Common Terminology Criteria for Adverse Events; [2]on Wechsler Memory scale III Word Lists, [3]RTOG, Radiation Therapy Oncology Group; [4]parameter fit from published data, [5]LENT-SOMA, late effects of normal tissues – subjective, objective, management and analytic.

its potential occurrence should be considered during treatment planning (Danesh-Meyer, 2010). Burman et al. (1991) developed an LKB model to estimate the probability of vision loss following irradiation of the optic nerves and chiasm. Based on the *gEUD* of each of these structures, three different predictions for blindness caused by the irradiation of the ipsi- or contralateral optic nerve or the chiasm were determined.

Brain and brain stem necrosis Necrosis of the brain or brain stem is a devastating late side effect that is essentially considered during treatment planning (Mayo et al., 2010b). To prevent necrosis at all costs, conservative dose constraints for the brain stem are usually applied and often given higher priority than target volume coverage. Moreover, it is difficult to differentiate radiation-induced necrosis from tumour progression, since both outcomes often occur in a similar pattern on MRI (Verma et al., 2013). Thus, confirmed incidences are rare so that hardly any NTCP models for modern treatment techniques exist. Bender (2012) developed an NTCP model for brain and brain stem necrosis five years after RT by combining outcome and dose data from case reports of several institutions summarised by Lawrence et al. (2010). Thus, the model is derived from different patient populations, fractionation schemes and treatment techniques. The maximum doses of the brain stem and brain-CTV were converted to EQD_2 to account for these differences.

Cataract requiring intervention Radiation to the crystalline lens of the eye may cause a cataract that commonly leads to vision impairment. As the lens is a radio-sensitive structure (Ferrufino-Ponce and Henderson, 2006), the maximum dose should be below 5 or 10 Gy (Henk et al., 1993; Lee et al., 2008; Piroth et al., 2009; Scoccianti et al., 2015). Although cataract surgery and replacement with artificial lenses are performed routinely, sparing during treatment planning is still the preferred solution, if possible. Burman et al. (1991) developed an LKB model to predict the probability of cataract requiring intervention based on *gEUD* of the lenses.

Delayed recall Neurocognitive sequelae are late side effects following cranial irradiation that may affect patients' QoL as described in detail in chapter 5. Gondi et al. (2012) assessed several neurocognitive functions in patients with LGG or benign brain tumours at 1.5 years following IMRT. Delayed recall on the WMS-WL was associated with $D40\%$ of the bilateral hippocampus.

Endocrine dysfunction Irradiation of the pituitary gland and hypothalamus may cause hormonal deficits such as impairment of growth hormone production, hyperprolactinemia and deficiency in gonadotrophin, thyroid-stimulating hormone and adrenocorticotropic hormone (Darzy and Shalet, 2009; Scoccianti et al., 2015). Dose constraints for the pituitary gland range between 50 and 60 Gy for adults (Pai et al., 2001; Lee et al., 2008). Endocrine functions outside the normal reference range (hyperprolactinemia, delayed thyroid-stimulating hormone response to thyroid-releasing hormone and panhypopituitarism) were modelled by De Marzi et al. (2015) based on

the dose to the pituitary gland for patients after combined treatment with PBT and 3D-CRT. They developed a logistic and an LKB model. As the predictions did not differ between both models, the NTCP values of the LKB model were considered in the present study.

Erythema Acute erythema as a radiation-induced reaction of the skin is comprehensively described in section 4.2. As NTCP models for modern XRT techniques are rare, a logistic model developed for PBT was applied (Dutz et al., 2019).

Ocular toxicity Irradiation of the lacrimal gland can affect tear production leading to reduced lubrication of the cornea and conjunctiva or dry eyes (Scoccianti et al., 2015). Fibrosis of the lacrimal gland with loss of tear production may be caused by doses of more than 57 Gy (Jeganathan et al., 2011). Batth et al. (2013) investigated ocular toxicity according to the RTOG scoring system. The degree of conjunctivitis, keratitis, corneal ulceration retinopathy and visual loss were the considered endpoint in patients with sinonasal tumours treated with IMRT. Acute ocular toxicity grade $\geq$ 2 could be predicted by the maximum dose to the ipsilateral lacrimal gland. As model coefficients were not given in this publication, the published dosimetric and outcome data were fitted in a logistic model in this thesis.

Ototoxicity Ototoxicity, i.e. hearing loss, tinnitus and/or vertigo, is a toxicity that occurs after irradiation of the ear and auditory nerve (Landier, 2016). Late sensorineural hearing loss has been described as a consequence of cochlear mean doses above 45 to 50 Gy (Pan et al., 2005; Lee et al., 2008; Bhandare et al., 2010). De Marzi et al. (2015) developed different logistic and LKB models to predict hearing loss for patients with tumours in the skull base after combined treatment with PBT and 3D-CRT based on the dose to the inner ear, cochlea and internal auditory canal. Hearing loss $>$ 15 dB at two contiguous test frequencies as well as tinnitus were considered as endpoints. Lee et al. (2015) investigated tinnitus grade $\geq$ 2 according to LENT-SOMA in head and neck cancer patients treated with IMRT. They developed a logistic and an LKB model based on the cochlear mean dose. For both side effects (hearing loss and tinnitus), the corresponding LKB model based on the dose to the cochlea was considered in the present study as the predictions of the logistic and the LKB model did not differ.

Temporal lobe injury Late temporal lobe necrosis is a severe side effect for patients treated with high doses for tumours of the skull base (Pehlivan et al., 2012). A study on patients with tumours located in the skull base treated with PBT investigated a relation of the occurrence of this side effect and dose to the temporal lobes. Although Pehlivan et al. (2012) observed a trend for grade 1 and grade 3 when *gEUD* exceeded 60 Gy(RBE), no NTCP model for temporal lobe injury could be developed. Another study investigated contrast-enhanced lesions on post-contrast T1-weighted MR images in patients for nasopharyngeal carcinoma treated with IMRT (Kong et al., 2016). They developed a logistic NTCP model for this endpoint at five years after IMRT based

on the maximum dose to the temporal lobe as predictor. As model coefficients were not given in the publication by Kong et al. (2016), the published dosimetric and outcome data were fitted in a logistic model in this thesis.

Evaluation of NTCP values

For each patient, NTCP values for all side effects were estimated for both treatment modalities, i.e. XRT and PBT. For some patients, the considered OAR was located within the tumour tissue, e.g. the pituitary for patients with pituitary adenoma. Thus, the NTCP value for the associated side effect could not be evaluated. Patient-individual differences $\Delta NTCP$ were assessed by

$$\Delta NTCP = NTCP_{\text{XRT}} - NTCP_{\text{PBT}}, \tag{6.5}$$

where $NTCP_{\text{XRT}}$ and $NTCP_{\text{PBT}}$ represent the $NTCP$ value based on the XRT and PBT treatment plan, respectively. Dose parameters were derived for the considered OARs based on the dose distribution on the planning CT scan.

For side effects with a given severity grade ≥ 2, a threshold of 10 percentage points was considered for $\Delta NTCP$ as suggested by Langendijk et al. (2018). However, not all side effects were physician-rated using scoring systems based on severity grades. For example, delayed recall was assessed using a neurocognitive test (Gondi et al., 2012). Due to the lack of further available recommendations for side effects not measured on a Likert scale and to ensure consistency, a threshold of 10 percentage points was also applied to all other side effects. A patient would finally be selected for PBT if $\Delta NTCP$ exceeds the threshold of 10 percentage points for at least one of all investigated side effects. For those cases where NTCP could not be predicted due to tumour invasion of the associated OAR, it was assumed that PBT is not more beneficial for the patient than XRT. For example, if one prediction was missing and the threshold was exceeded for at least one other side effect, the patients would be selected for PBT. If one prediction was missing and the threshold was not exceeded for any other side effect, the patients would be selected for XRT.

6.3.2 Results

The NTCP predictions of all investigated side effects and both modalities are given in table 6.6. For temporal lobe injury (Kong et al., 2016), contralateral hearing loss (De Marzi et al., 2015) as well as brain and brain stem necrosis (Bender, 2012), the maximum NTCP value for both RT modalities were below 5 %. Therefore, these side effects were not considered any further. For the remaining side effects, all $\Delta NTCP$ values and a threshold of 10 percentage points are shown in figure 6.5. The NTCP values for both techniques covered a wide range in some models, especially those predicting side effects after irradiation of ipsilateral organs such as cataract and ocular toxicity as well as tinnitus and hearing loss. But also the NTCP values for delayed recall, acute

Table 6.6: NTCP predictions for different side effects following cranial RT. Mean and standard deviations are given in the first row, median and range are shown in the second row. P-values are derived from Wilcoxon signed rank tests. Number of patients, for which $\Delta NTCP$ exceeds a threshold of 10 percentage points (pp) is given in the right column.

Side effect	$NTCP_{PBT}$ in %	$NTCP_{XRT}$ in %	$\Delta NTCP$ in percentage point	p-value	Patients with $\Delta NTCP \geq 10$ pp
Alopecia	60.2 ± 25.0	63.5 ± 20.8	3.3 ± 8.5	0.001	21/92
Dutz et al. (2019)	63.2 (20.9 – 96.8)	60.0 (31.4 – 96.9)	3.0 (-19.8 – 21.0)		
Blindness (chiasm)	2.2 ± 3.9	2.8 ± 4.1	0.5 ± 2.0	< 0.001	1/91
Burman et al. (1991)	0.0 (0.0 – 13.3)	0.2 (0.0 – 17.5)	0.0 (-3.3 – 12.3)		
Blindness (ON contra)	0.1 ± 0.8	0.2 ± 1.3	0.1 ± 1.1	< 0.001	1/92
Burman et al. (1991)	0.0 (0.0 – 7.7)	0.0 (0.0 – 10.3)	0.0 (-0.6 – 10.2)		
Blindness (ON ipsi)	1.5 ± 3.2	1.8 ± 3.3	0.3 ± 1.1	< 0.001	0/91
Burman et al. (1991)	0.0 (0.0 – 15.1)	0.0 (0.0 – 12.9)	0.0 (-4.1 – 4.3)		
Cataract (contra)	0.1 ± 0.4	4.1 ± 15.9	4.0 ± 15.6	< 0.001	6/92
Burman et al. (1991)	0.0 (0.0 – 3.4)	0.1 (0.0 – 99.7)	0.1 (0.0 – 99.4)		
Cataract (ipsi)	5.7 ± 20.4	21.7 ± 35.4	16.0 ± 35.0	< 0.001	24/91
Burman et al. (1991)	0.0 (0.0 – 100.0)	0.3 (0.0 – 100.0)	0.2 (-85.8 – 99.8)		
Delayed recall	31.6 ± 38.3	65.3 ± 34.7	33.7 ± 32.4	< 0.001	51/78
Gondi et al. (2012)	4.9 (3.2 – 100.0)	77.1 (3.7 – 100.0)	24.8 (-29.1 – 93.8)		
Endocrine dysfunction	1.3 ± 3.1	1.8 ± 3.7	0.4 ± 1.4	< 0.001	0/87
De Marzi et al. (2015)	0.0 (0.0 – 13.8)	0.0 (0.0 – 18.1)	0.0 (-2.0 – 8.6)		
Erythema	34.0 ± 21.3	34.5 ± 21.5	0.5 ± 8.1	0.63	8/92
Dutz et al. (2019)	23.0 (17.7 – 99.0)	23.6 (17.7 – 95.9)	0.0 (-32.9 – 24.6)		
Hearing loss (contra)*	0.0 ± 0.0	0.0 ± 0.4	0.0 ± 0.4	< 0.001	0/92
De Marzi et al. (2015)	0.0 (0.0 – 0.1)	0.0 (0.0 – 3.6)	0.0 (-0.1 – 3.5)		
Hearing loss (ipsi)	4.7 ± 12.1	6.4 ± 15.5	1.7 ± 11.6	< 0.001	7/92
De Marzi et al. (2015)	0.0 (0.0 – 61.8)	0.0 (0.0 – 67.0)	0.0 (-41.6 – 65.8)		
Necrosis (brain stem)*	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.007	0/92
Bender (2012)	0.0 (0.0 – 0.2)	0.0 (0.0 – 0.1)	0.0 (-0.1 – 0.0)		
Necrosis (brain)*	0.1 ± 0.1	0.1 ± 0.1	0.0 ± 0.0	0.095	0/92
Bender (2012)	0.1 (0.0 – 0.3)	0.1 (0.0 – 0.3)	0.0 (-0.1 – 0.1)		
Ocular toxicity (ipsi)	30.4 ± 40.6	36.0 ± 40.3	5.6 ± 26.6	< 0.001	24/91
Batth et al. (2013)	1.1 (0.6 – 99.9)	9.6 (0.6 – 99.9)	1.6 (-69.6 – 88.8)		
Temporal lobe injury (contra)*	0.1 ± 0.1	0.1 ± 0.2	0.0 ± 0.0	< 0.001	0/92
Kong et al. (2016)	0.0 (0.0 – 0.9)	0.0 (0.0 – 1.0)	0.0 (-0.1 – 0.3)		
Temporal lobe injury (ipsi)*	0.3 ± 0.3	0.4 ± 0.4	0.0 ± 0.2	0.044	0/80
Kong et al. (2016)	0.2 (0.0 – 1.4)	0.2 (0.0 – 1.6)	0.0 (-0.2 – 0.7)		
Tinnitus (contra)	0.7 ± 2.1	2.6 ± 5.3	1.9 ± 3.8	< 0.001	3/92
Lee et al. (2015)	0.2 (0.2 – 16.9)	0.5 (0.2 – 39.6)	0.3 (-1.1 – 22.7)		
Tinnitus (ipsi)	14.5 ± 23.0	17.2 ± 24.2	2.6 ± 10.6	< 0.001	13/92
Lee et al. (2015)	0.3 (0.2 – 77.1)	3.8 (0.2 – 78.9)	0.4 (-49.8 – 49.6)		

Abbreviations: ON, optic nerve; ipsi, ipsilateral; contra, contralateral. *excluded from further analysis as maximum NTCP for both modalities <5 %

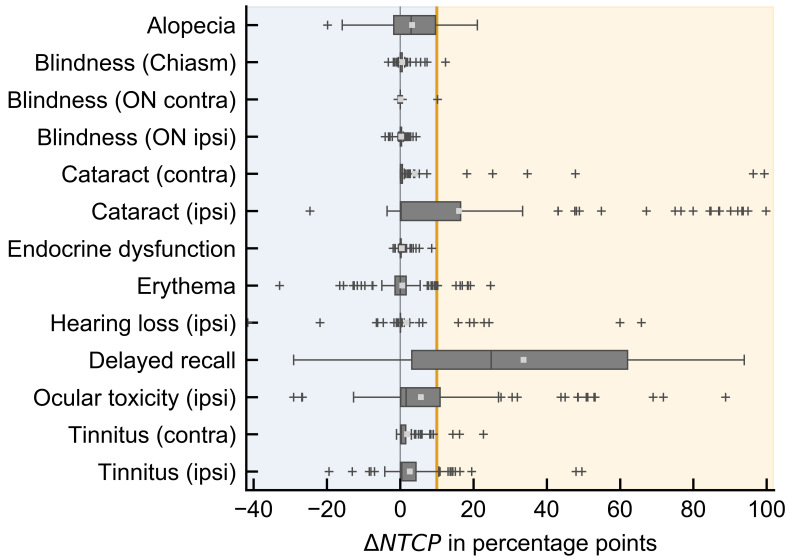


Figure 6.5: NTCP differences between XRT and PBT plans for different side effects. A threshold of 10 percentage points is indicated as orange line. ON, optic nerve; ipsi, ipsilateral; contra, contralateral.

alopecia and erythema showed a large spread. In general, low probabilities of developing blindness (highest mean value: 1.8 % with XRT) and endocrine dysfunction (mean value: 1.8 % with XRT) were observed. For the majority of the investigated models, the differences between XRT and PBT predictions were statistically significant, except for acute erythema and brain necrosis. The significance only indicates that NTCP predictions differ systematically between both modalities for most side effects. However, the extent of these differences is not evaluated and should be therefore characterised by the use of threshold values.

Of all NTCP predictions, the highest median NTCP value was estimated for delayed recall for XRT plans (77.1 %). For PBT plans, acute alopecia had the highest median NTCP (63.2 %) among all investigated side effects. For almost all side effects, the median difference between $NTCP_{\mathrm{PBT}}$ and $NTCP_{\mathrm{XRT}}$ remained below 5 percentage points except for delayed recall (Gondi et al., 2012). For this side effect, the median $\Delta NTCP$ was estimated to 24.8 percentage points. Applying a threshold of 10 percentage points, 51 of 78 patients (65.4 %) would have been selected for PBT based on this side effect. In 14 patients the tumour completely affected the ipsilateral hippocampus. Hence, it could neither be delineated, nor could the bilateral hippocampi be evaluated. For both ipsilateral ocular toxicity and cataract, 24 of 91 patients (26.4 %) exceeded the threshold of 10 percentage points. For ipsilateral hearing loss, contralateral cataract and acute erythema, only 6 to 8 of 92 patients would have been selected (6.5 – 8.7 %). No patients would have been selected for PBT based on the predictions for blindness (Burman et al., 1991) or endocrine dysfunction (De Marzi et al., 2015) using a threshold of 10 percentage points.

All side effects for which $\Delta NTCP$ of more than 10 percentage points was predicted in at least five patients are depicted in figure 6.6. Here, the absolute NTCP estimates and different pat-

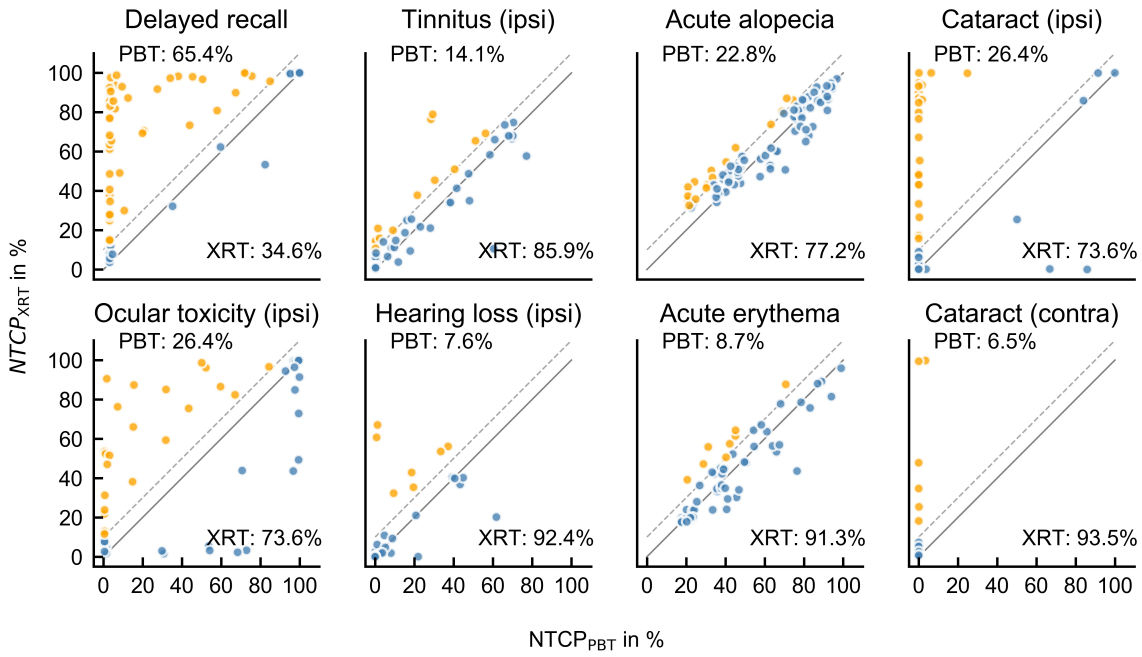


Figure 6.6: NTCP predictions for XRT and PBT plans for selected side effects. The solid line represents equal NTCP values for both treatment modalities. A threshold of 10 percentage points is indicated as a dashed line. The percentage of patients selected for each treatment based on this threshold is given.

terns can be observed. For some side effect, the NTCP estimates were generally low for both treatment techniques (e.g. contralateral tinnitus: $NTCP < 20\,\%$ for almost all patients). For other side effects, the estimates for both techniques were higher for all patients and both techniques (e.g. acute alopecia: $NTCP > 20\,\%$). For cataract, the NTCP predictions for PBT were nearly zero for the majority of patients, while a greater spread was estimated for XRT. This fact leads to larger NTCP differences so that 26.4 % and 6.5 % of the patients exceeded the 10 percentage points threshold for ipsi- or contralateral cataract, respectively.

For the majority of patients, the probability of developing one (32 patients) or two of the investigated side effects (30 patients) was reduced by more than 10 percentage points using PBT instead of XRT, see table 6.7. In twelve patients, $\Delta NTCP$ did not exceed the required threshold for any of the investigated side effects.

Table 6.7: Number of patients and side effects with $\Delta NTCP$ exceeding a threshold of 10 percentage points. For those cases where NTCP could not be predicted due to tumour invasion of the associated OAR, it was assumed that PBT is not more beneficial than XRT.

Side effects with $\Delta NTCP \geq$ 10 percentage points	**0**	**1**	**2**	**3**	**$\geq$4**
Patients (%)	12 (13.0)	32 (34.8)	30 (32.6)	9 (9.8)	9 (9.8)

The NTCP differences between both treatment modalities for the considered side effects and each patient are presented in appendix D figure D.3. Appendix D figure D.4 shows the side effects for which $\Delta NTCP$ exceeds a threshold of 10 percentage points for all patients.

For the two example patients, already discussed in section 6.2 and presented in table 6.4 and figure 6.3, $\Delta NTCP$ estimates for the investigated side effects and the corresponding dosimetric parameters are displayed in figure 6.7 for each NTCP model.

For patient 1, for only one side effect, namely delayed recall, the NTCP reduction using PBT instead of XRT exceeded the threshold of 10 percentage points. All other $\Delta NTCP$ estimates remained below this threshold. The probability of developing acute alopecia and erythema were

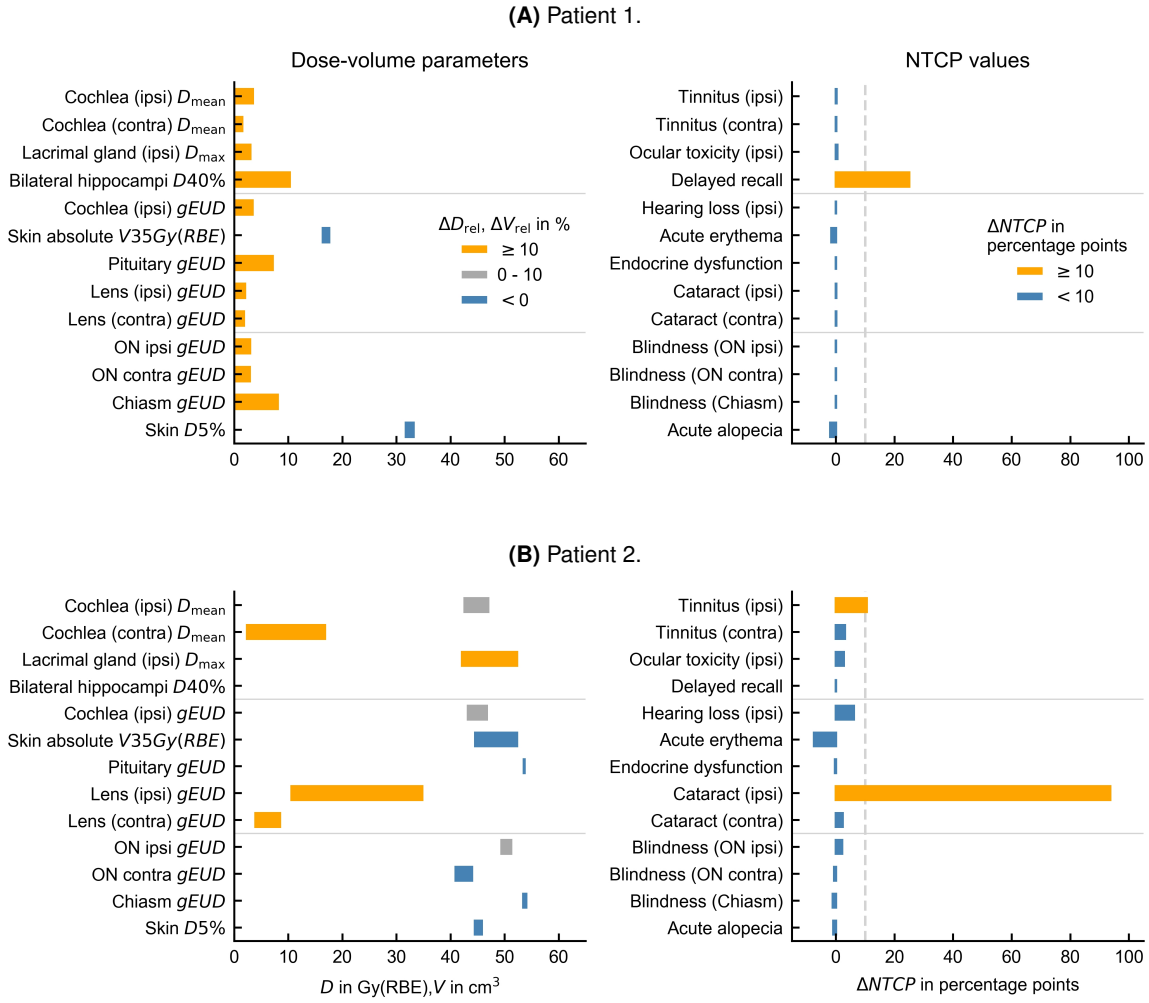


Figure 6.7: Differences of DVH parameters between the XRT and PBT plan (left panel) and associated NTCP values for selected side effects (right panel) for two patients. The dashed line indicates a threshold of 10 percentage points. DVH parameters with a relative difference of more than 10 % and side effects with $\Delta NTCP$ exceeding the threshold of 10 percentage points are shaded in orange. ON, optic nerve; ipsi, ipsilateral; contra, contralateral.

even slightly reduced using XRT. Although the relative differences in dosimetric parameters were more than 10 % (except for the skin), they could not be translated into a potential clinical benefit for the majority of side effects using PBT. The absolute doses for these parameters were relatively low for both treatments and are thus located in the flat initial region of the NTCP curves. As a result, the NTCP values of both therapies hardly differ. For patient 2, the substantially reduced dose to the lenses resulted in a large reduction of the probability of ipsilateral cataract induction when treated with PBT instead of XRT ($\Delta NTCP > 80\,\%$). Additionally, the potential NTCP difference for ipsilateral tinnitus exceeded the threshold of 10 percentage points. Remarkably, that the relative and absolute dose difference for the mean dose to the ipsilateral cochlea was quite small (4.3 Gy(RBE) and 9.2 %). However, the small dose difference could be converted into an NTCP reduction of more than 10 percentage points using PBT instead of XRT. In comparison, the dose reduction in the contralateral side was substantially greater for PBT (14.3Gy(RBE) and 83.6 %). Nevertheless, the NTCP difference for contralateral tinnitus remained below the threshold of 10 percentage points. For this patient, the probability of developing acute erythema and alopecia would be slightly increased using PBT instead of XRT.

Differences in NTCP reduction with regard to tumour location

Median NTCP predictions for both treatment modalities as well as $\Delta NTCP$ were analysed separately for different tumour locations, see appendix D table D.3. For the majority of side effects, the median $\Delta NTCP$ estimates remained below 3 percentage points for the investigated tumour locations. However, for delayed recall, a much higher median $\Delta NTCP$ compared to the entire cohort (24.8 percentage point) could be observed for patients with tumours located in the parietal (43.9 percentage point) or temporal lobe (45.4 percentage point). For patients with a tumour located in the frontal lobe, this estimated $\Delta NTCP$ was comparable with the entire cohort (25.0 percentage point), while for patients with a tumour located in more than one lobe, lower $\Delta NTCP$ values were observed (10.4 percentage point). For tumours in the parietal lobe, median $\Delta NTCP$ for alopecia (11.2 percentage point) and erythema (18.4 percentage point) were increased compared to the entire cohort (3.0 and 0.0 percentage point, respectively).

6.3.3 Discussion

For the majority of patients and a variety of side effects, NTCP predictions were higher for XRT than for PBT as expected, but $\Delta NTCP$ did often not exceed a threshold of 10 percentage points for an individual side effect. However, based on the model selection in this study, a majority of patients (87.0 %) would still have been selected for PBT if this decision would be based on a threshold of 10 percentage points for at least one side effect. One reason for this effect is that a variety of different side effects have been considered in this study. Furthermore, individual side effects had a high impact on the selection process, namely delayed recall (Gondi et al., 2012). 51 of 78 patients would have been selected for PBT only because of this side effect. Moreover, no

ranking or weighting of the clinical severity between different side effects was done. For example, permanent ocular toxicity, ranging from conjunctivitis, keratitis, corneal ulcer, retinopathy or detachment to loss of vision, might be more severe than cataract because it is possible to replace the crystalline lens. If such a ranking of side effects should be considered for a model-based selection, it has to be established by experienced radiation oncologists.

Although a variety of NTCP models were analysed in this study covering a wide range of potential side effects that may occur following cranial RT, there may be even more side effects, for that no NTCP model exists so far or the underlying dose-effect relationship is not yet fully understood. Especially for neurocognitive dysfunction, more reliable NTCP models than the model for delayed recall by Gondi et al. (2012) have to be developed as it turned out to overestimate the actual incidence as explained in detail below (Jaspers et al., 2019). This important side effect may be one of the decisive side effects for the selection of patients for PBT as it adversely affects patients' QoL. In the present study, 16 patients (17.4 %) would have been selected for PBT solely based on delayed recall. In the other patients with $\Delta NTCP$ for delayed recall above 10 percentage points, at least one more side effect would have been decisive for treatment with PBT.

Influence of tumour- and treatment-specific parameters In general, NTCP differences for specific tumour locations did not differ from those of the entire cohort. In patients with a tumour in multiple lobes, $\Delta NTCP$ for delayed recall was at least strongly pronounced. Patients with a tumour in the parietal or temporal lobe showed particularly strong reductions in delayed recall compared to the entire cohort. These patients might particularly benefit from PBT since neurocognitive dysfunction is a major side effect. However, as explained in the dose-comparison study (see section 6.2), a dedicated localisation group may include many different specific tumour sites and therefore cannot provide a general basis for selection for PBT. Thus, it is still important to conduct a patient-specific analysis.

As described in the dose-comparison study, the relative dose reduction in OARs achieved by PBT is particularly prominent in contralateral structures. However, as the absolute doses in the contralateral structures are fairly low for both techniques, both NTCP estimates also remain low, leading to small $\Delta NTCP$ values. For ipsilateral structures, the NTCP differences were often higher compared to contralateral structures although the relative dose differences were smaller. Thus, a greater portion of the patients would be selected for PBT due to the ipsilateral side effects rather than the contralateral. This effect was observed for the bilateral side effects hearing loss, tinnitus as well as cataract.

For delayed recall, a correlation between increasing $\Delta NTCP$ values and increasing tumour volume (CTV) (Spearman ρ = 0.78, $p < 0.001$) as well as increasing prescribed dose was observed (Spearman ρ = 0.56, $p < 0.001$). This suggests that patients with large tumours and a high prescription dose may benefit from PBT to a larger extent with regard to this side effect than

others. For all other investigated side effects, such correlations (with Spearman $|\rho| > 0.5$) could be observed for neither tumour volume nor prescribed dose.

When comparing $\Delta NTCP$ predictions between the subgroups treated with PBS and DS, no significant differences were observed for alopecia, blindness (chiasm, contralateral optic nerve) and delayed recall. For the other side effects, the benefit from using PBT instead of XRT was increased in the group treated with PBS compared to DS for cataract, ocular toxicity, endocrine dysfunction, ipsilateral hearing loss and tinnitus. Significantly more patients would be selected for PBT based on a threshold of 10 percentage points for cataract (ipsilateral $p < 0.001$ and contralateral, $p = 0.026$) and ocular toxicity ($p = 0.031$) in the group treated with PBS compared to a DS treatment. However, these side effects depend largely on tumour size and location as well as the prescribed dose. Similarly to the dose-comparison study, a comprehensive comparison between PBT techniques requires a PBS and DS plan for each patient. This investigation was out of the scope of this study.

NTCP model for delayed recall Of all investigated NTCP models, the model for delayed recall on the WMS-WL at 1.5 years following IMRT (Gondi et al., 2012) predicted the largest median difference between $NTCP_{\text{PBT}}$ and $NTCP_{\text{XRT}}$. Based on this side effect, 65.4 % of the patients would have been selected for PBT using a $\Delta NTCP$ threshold of 10 percentage points. A recent study investigated the performance of this model on an external cohort of LGG patients treated with XRT (Jaspers et al., 2019). They observed that the model predictions overestimated the actual neurocognitive outcome measured by the Rey Verbal Auditory Learning test delayed recall (AVLT-DR). For 22 of 29 patients, a probability for neurocognitive dysfunction of more than 99 % was predicted. However, only 7 patients were found to have a decline. They explained this by the fact that the model was developed for a narrow dose range, but was extrapolated to a larger dose range during validation. Figure 6.8 shows the NTCP model for delayed recall and the dose range for that it was developed on. As the steep part of the NTCP curve is located in a narrow dose range up to 30 Gy, dose differences between both treatment techniques in this dose range lead to large NTCP differences. In contrast to this model, the NTCP model for tinnitus grade ≥ 2 is also depicted in figure 6.8 (Lee et al., 2015). This model was developed within a broad dose range up to 60 Gy. The slope of the NTCP curve is much shallower than of the NTCP model for delayed recall. Special care should be taken when using NTCP models outside the dose range in which they were developed.

Selection of appropriate thresholds In the present study, a $\Delta NTCP$ threshold of 10 percentage points above which a patient would be selected for PBT was applied to all NTCP models. Langendijk et al. (2018) proposed this threshold for side effects grade 2 that may occur after head and neck RT. As some NTCP models included in this study predicted side effects of grade ≥ 2 and no further suggestions had been available for the NTCP model predicting delayed recall, this specific threshold was chosen. Delayed recall was not measured on a Likert-scale as it is

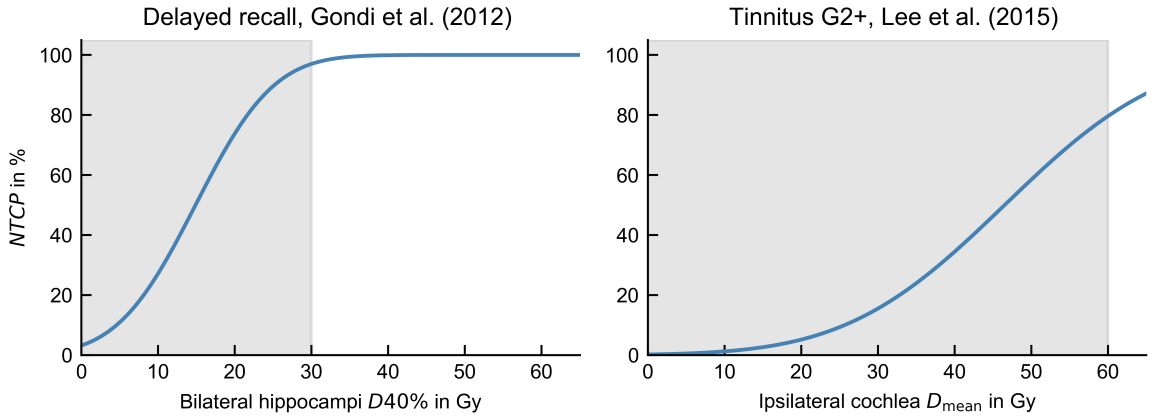


Figure 6.8: Applied NTCP models for delayed recall (Gondi et al., 2012) and tinnitus (Lee et al., 2015). Dose ranges for which the respective NTCP models were developed are shaded in grey.

the case for other measures such as CTCAE. As the number of patients selected for PBT based on delayed recall was 65.4 %, the applied threshold may be disputable. Figure 6.9 shows the patient selection distribution for different $\Delta NTCP$ thresholds for delayed recall. An increase of the threshold to 20 or 30 percentage points would reduce the number of patients selected for PBT to about 57.7 % and 43.6 %, respectively. This demonstrates that a comprehensive analysis of the potential NTCP reduction should be conducted before the implementation of the model-based approach in clinical practice. Similar plots are depicted for the remaining investigated NTCP models in appendix D figure D.5. As NTCP models were derived from physician-rated side effects recorded on different scoring systems (CTCAE, RTOG or LENT-SOMA) it should be investigated whether the threshold suggestions based on the CTCAE scale are also reasonable for other scoring systems.

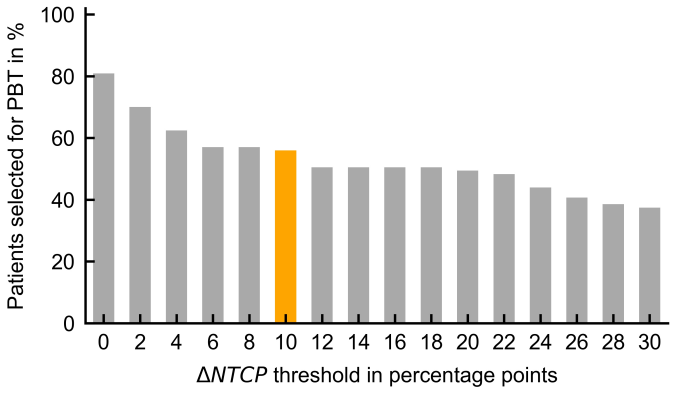


Figure 6.9: Patient selection for PBT based on the NTCP model for delayed recall for different thresholds (Gondi et al., 2012). The applied threshold of 10 percentage points is shaded in orange.

Time-dependence of the onset of side effects The present study included NTCP models of both acute and late side effects. Figure 6.10 shows the time spans for the occurrence of these side effects that were endpoints of the considered NTCP models. For patients with a short life expectancy (e.g. HGG), acute side effects may be more important to maintain QoL as long as possible than side effects that have an onset after several years. In contrast, patients with a long life expectancy (e.g. benign tumours) may experience late side effects as more burdensome than potential transient early side effects. The curves in figure 6.10 symbolise exemplarily the de- and increasing importance of different side effects for both patient groups. Consequently, the different NTCP predictions could be weighted with a time-dependent weighting factor with regard to the life expectancy. This could alter patient selection according to the model-based approach. The development of such weight functions for different patient groups should be done in close collaboration with physicians.

Re-irradiated patients The study included six patients that received re-irradiation with PBT. For these patients, PBT may have been chosen because the normal tissue has already been affected by the previous irradiation assuming that the steep dose gradient in PBT could better spare this exposed tissue. However, since the model-based approach is intended to be used for all patients without standard indications for PBT, re-irradiated patients were also considered in this study. Results of the analysis for all patients without re-irradiation are presented in appendix D figures D.6 and D.7. The proportion of patients selected and the other conclusions barely change. In these six patients, the first irradiation was at least two years, but in most patients even five years prior to PBT, so that early reactions had completely subsided. It has been shown, that the skin regains the primary tolerance threshold for acute skin reactions. In studies on re-irradiated patients with head and neck cancer, breast cancer and others, acute skin reactions were within the same range as those at first irradiation (Nieder and Langendijk, 2016).

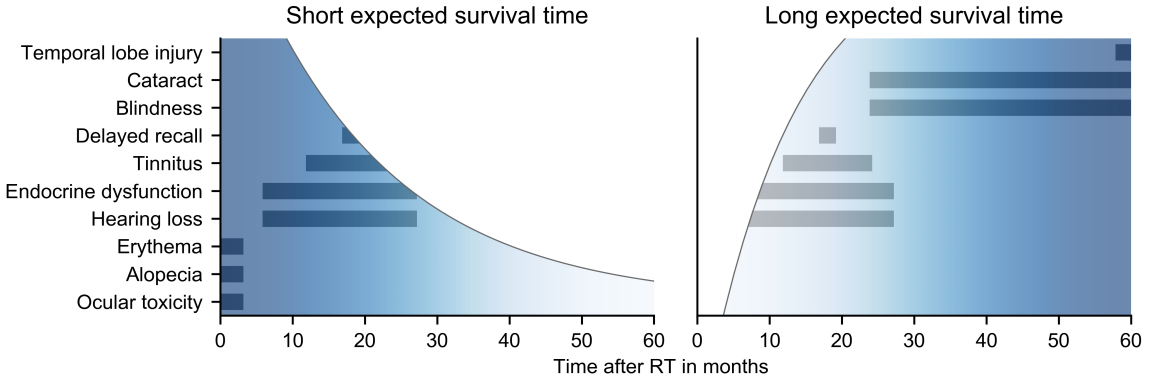


Figure 6.10: Importance of side effects for patients with different life expectancies. Observation time spans of various side effects underlying the investigated NTCP models are displayed as bars. The curves symbolise exemplary weight functions representing the de- and increasing importance of certain side effects for both patient groups.

Thus, it seems appropriate to use models for early alopecia and erythema even for re-irradiated patients. Late side effects associated with re-irradiation of the brain are difficult to assess, as not even data from animal experiments are available. Moreover, depending on the blood supply of different brain areas, the toxicity risk may also differ. It is generally assumed that late side effects will persist for many years after the treatment. However, there is limited knowledge about re-irradiation tolerance for several important late effects, including neurocognitive or endocrine dysfunction (Nieder and Langendijk, 2016). To predict the complication risk of an individual patient for choosing an optimal treatment strategy, validated NTCP models for re-irradiation would be required. However, developing such models will be even more difficult than those for initial RT treatment. First, fewer patients may be eligible compared to first-course treatment, which limits the cohort size. Furthermore, additional parameters are needed to characterise the pre-irradiation, such as the dose of the initial irradiation, the time interval between both RT courses and concomitant systemic therapies (Eekers et al., 2016).

Radiation-induced brain necrosis In this study, the probability of the induction of brain or brain stem necrosis was estimated to be comparatively low with a maximum NTCP value of 0.3 % for both techniques based on the model by Bender (2012), which was derived from XRT data. Development of NTCP models for radiation-induced necrosis is demanding as a distinct differentiation of necrotic tissue from tumour progress is challenging as discussed in section 4.3.2.

Whether the models developed based on XRT data can also be utilised in PBT, is questionable. Although radiation-induced necrosis is a rare side effect, reports of brain stem necrosis after PBT indicate possible biological differences between XRT and PBT (Haas-Kogan et al., 2018). For proton beams, both LET and RBE are variable and increase in the distal edge of the spread-out Bragg-Peak leading to enhanced biological effects (Paganetti, 2014; Lühr et al., 2018). There is a growing interest in investigating the influence of LET and RBE on the treatment to estimate the development of contrast-enhancing lesions that may evolve into necrotic brain regions (Eulitz et al., 2019a; Eulitz et al., 2019b). Since brain necrosis is such a severe side effect, it should definitively be considered when applying the model-based approach, as soon as these models exist and are validated. Due to the severity of this side effect, a rather low threshold value should be applied.

The effect of increased biological effectiveness could of course also affect other side effects besides brain necrosis. This would require the validation of existing XRT models on PBT patients. However, this was not feasible in the context of the study. The published photon-derived models predicted side effects that on the one hand were not recorded in the PBT trials used in this thesis (DRKS00004384, 2012; DRKS00007670, 2015; DRKS00008569, 2015; NCT02824731, 2016) or on the other hand the follow-up time within these PBT trials was still too short. And vice-versa, for the applied proton-derived models, there was no outcome data of patients treated with XRT available for validation.

Competing side effects So far, the model-based approach assumes that the use of PBT instead of XRT will reduce the occurrence of side effects. This assumption is based on dosimetric plan comparisons and only the extent of the expected reduction is regarded for a patient to be treated with PBT. However, it should be discussed whether PBT might not also intensify some side effects. This effect may even be enhanced when considering LET and RBE variations of PBT. Negative threshold values in favour of XRT could be defined. Consequently, competing side effects have to be considered, e.g. PBT may reduce the probability of one side effect while potentially increasing that of another side effect.

Treatment planning approaches For both, PBT and XRT plans, a patient individual beam arrangement was used according to clinical practice. Besides the selection of the beam arrangements, the objectives, as well as their weightings during treatment plan optimisation, have an important impact on the final dose distribution. The different prioritisation of several OARs depending on the severity of the associated side effects may lead to bias. In both treatment modalities, a high priority – even higher than target coverage – was given to the avoidance of high-dose areas in the brain stem to prevent necrosis. Hence, almost no difference between PBT and XRT was seen in the NTCP values for brain stem necrosis with the applied NTCP models. Differences in the prediction of side effects are most likely to be achieved for those side effects for which the severity of their occurrence is of minor importance compared to the target coverage, e.g. cataract. This also applies to OARs for which clinical dose constraints have not yet been specified and which are therefore not yet considered in treatment planning. So far, treatment plan optimisation mainly includes physical dose parameters focusing on OARs for which well-established dose constraints exist. Furthermore, most often only these structures are delineated in clinical routine.

With the availability of intensity-modulated RT techniques, highly conformal dose distributions can be achieved that can be much more adapted compared to 3D-CRT techniques. One approach to account for this development is biological treatment planning. In addition to physical optimisation parameters, biological optimisation parameters are included in the objective function. The most common biological optimisation parameters are based on *gEUD* depending on the tissue-specific parameter *a* (Niemierko, 2006). It has been shown that intensity-modulated treatment plans, generated by biological optimisation parameters, lead to an improved sparing of normal tissue compared to objective functions with conventional physical dose-volume parameters (Qi et al., 2009). This seems to be a promising approach to reduce side effects. Nevertheless, caution is required when selecting adequate models and tissue-specific model parameters, as valid models are not yet available for all tumour types and side effects. Moreover, model uncertainties are not considered during biological optimisation (Qi et al., 2009). Another approach, which goes one step further than just using *gEUD*, directly incorporates multivariable NTCP models in the objective function (Kierkels et al., 2014). Provided that reliable NTCP models are available, the identification of optimal planning objectives could be simplified. Such personalised

and automated optimisation could make the creation of comparison plans more efficient (Wu et al., 2003). This, in turn, could lower the inhibition threshold for introducing the model-based approach in clinical practice, as *in silico* plan comparison is very time-consuming.

Limitations Since the comparison of the NTCP values is based on the underlying treatment plans, all limitations of the treatment plan comparison also apply here, see section 6.2. Additionally, there are further limitations regarding the NTCP models and their selection described below.

All dosimetric parameters and NTCP predictions were calculated based on the planned dose distribution and nominal model coefficients. The uncertainties of each model coefficient as well as differences of nominal dose to delivered dose were not considered in this study. However, model and dose uncertainties may influence the accuracy of the selection process. Since treatment decision is taken before treatment, model-based patient selection is based on the planned dose rather than the accumulated delivered dose. The planning CT is only a single snapshot that is not representative of the anatomical situation during every treatment fraction. For OARs with small volumes, such as lens, cochlea or lacrimal gland, slight positioning errors can lead to deviations of the delivered dose from the planned dose, which in turn, may alter the probability of associated side effects as estimated from the planned dose distribution. To evaluate dose uncertainties prior to PBT, robust treatment plan optimisation or analyses can be performed. Both methods consider positioning as well as calculation uncertainties caused by the conversion of CT numbers to mass stopping power.

Bijman et al. (2017) evaluated the model-based patient selection for three different NTCP models on a cohort of 78 patients with oropharyngeal cancer. They calculated nominal NTCP differences for an IMRT and a PBT plan based on the initial planning dose distributions and nominal model coefficients. Additionally, the effects of uncertainties in the model coefficients and differences from delivered to planned dose were investigated by a sampling of model coefficients and doses from their probability distribution. Selection accuracy defined as the fraction of consistent selections in both the nominal and sample scenarios were about 70 % for NTCP model uncertainties only. It further decreased by excessive dose uncertainties. For patients with head and neck cancer, both tumour and OARs may change remarkably during treatment due to tissue responses such as tumour shrinkage or tissue swelling (Barker et al., 2004). Therefore, differences between planned and delivered doses are common in head and neck patients, which adaptive RT is intended to compensate. Brain tumour patients may be less prone to these variations. However, tissue changes close to the resection cavity might be considered for post-operative patients (Kim and Lim, 2013).

Besides dose and model coefficient uncertainties, the NTCP models developed at other institutions are subject to further uncertainties such as differences in treatment techniques, delineation guidelines, scoring of side effects or patient population. The patient population may differ with regard to tumour histology and location. In this treatment planning study, a VMAT technique was

applied as a state-of-the-art XRT technique. However, the investigated NTCP models are derived from different techniques including 3D-CRT or IMRT, as the available treatment techniques changed rapidly during the last decades. Different XRT techniques may alter the predictions of the NTCP models (Christianen et al., 2012).

Differences in the treatment fractionation schemes were corrected using EQD_2 values, for example in the model predicting brain necrosis by Bender (2012). Different scoring systems are used at different centres to assess radiation-induced side effects (e.g. CTCAE, RTOG or LENT-SOMA). Severity grades may vary depending on the systems used (Höller et al., 2003). Moreover, the interobserver variability among clinicians can be rather high, especially when monitoring subjective symptoms such as pain, nausea or fatigue (Brundage et al., 1993; Neben-Wittich et al., 2011).

Another important uncertainty arises from variations in the delineation of OARs. Associated dosimetric parameters are affected, which in turn modifies the predictions of the NTCP models (Brouwer et al., 2014). Although standardised delineation guidelines were used in this study, it remains unclear which guidelines have been used to contour the OARs included in the applied published models. The issue of potentially different structures is likely to arise if the models are not generated from one cohort but from several data sets, such as the model for brain necrosis (Bender, 2012). However, such studies cover a large variability, which nevertheless occurs in everyday clinical practice. To reduce contouring variability, consistent use of delineation guidelines and atlases is essential. Additionally, the implementation of auto-contouring tools could be considered (Vinod et al., 2016). These tools are currently being further improved, e.g. by deep-learning-based segmentation approaches (van der Heyden et al., 2019; van Dijk et al., 2020). Besides reducing the delineation time, auto-contouring can also facilitate the segmentation of additional OARs. These could be used in NTCP modelling to investigate unknown relationships between dose to these structures and side effects such as neurocognition.

In everyday clinical practice, existing NTCP models have to be used, since it is not feasible for an institute to generate its own models for all possible side effects, patient subgroups and histologies, treatment techniques, concomitant therapies, uniform contouring, fractionation schemes and available long-term follow-up, etc. However, the existing models should be validated on centre-specific data, if possible. This requires a continuous long-term scoring of side effects, with the same tests and scoring systems that were used for model creating. Prior to the implementation of a model-based patient selection for PBT, a comprehensive analysis of side effects and existing models is necessary.

In this thesis, available NTCP models were evaluated for a variety of side effects, although they did not reach the highest level of the TRIPOD classification, see appendix A table A.3. This is because validated models for adverse effects after cranial RT with modern techniques are scarce. It was not possible to validate either the XRT or the PBT models investigated. For example, the clinical outcome data of the patients enrolled studies used for this thesis were recorded by different scoring systems than in the investigated published models. Moreover, the

follow-up time was too short to detect late side effects. No clinical outcome data were available for similar cohorts treated with modern XRT techniques. Although NTCP differences rather than absolute values were evaluated in this study, the presented results provide only an estimate of the potential probability of side effects and should therefore be interpreted with caution.

Induction of secondary malignancies is a rare complication of RT with long latency between treatment and onset. The probability of developing a radiation-induced secondary tumour depends on the type of tissue, the dose and irradiated volume as well as the age of the patient (Prasad and Haas-Kogan, 2009; Schneider, 2011). As children are considered more radiosensitive than adults while having a longer life expectancy, retrospective analyses were mainly conducted in paediatric populations treated with outdated RT techniques due to the long time between irradiation and onset of the secondary tumour. One of these studies investigated secondary malignancies in patients treated for childhood cancer. Interestingly, one-third of the secondary tumours were located in areas receiving very low doses $< 2.5\,\mathrm{Gy}$ (Diallo et al., 2009). This is of particular concern for modern intensity-modulated XRT techniques, as they deliver relatively high distant peripheral doses compared to 3D-CRT or PBT (Klein et al., 2006). For adults, the data situation is even more limited. Available data are derived from atomic bomb survivors, nuclear industry workers, clean-up workers after the Chernobyl nuclear accident or adults after repeated CT imaging (Prasad and Haas-Kogan, 2009). Models based on these data were developed mainly for radiation protection and are based on lower doses than those applied during RT treatment and should therefore not be applied for risk estimates in RT. Since fractionation affects carcinogenesis, better models include cell repopulation between dose fractions. However, the uncertainties are still very large for most tissues as the underlying processes are not well understood (Schneider, 2011). As RT is part of the standard treatment of patients with brain tumours, it is very challenging to find cohorts that compare irradiated and non-irradiated adults (Prasad and Haas-Kogan, 2009). Since there are practically no models appropriate for secondary tumours after cranial irradiation, they were not evaluated in this study. Moreover, for HGG patients, secondary malignancies may play a minor role due to the limited life expectancy.

The patients included in this study cover a wide range of tumour histologies and locations. However, some tumour histologies (e.g. high-grade meningioma) and tumour locations (e.g. occipital) were underrepresented. For patients with a tumour in the skull base, additional OARs may be relevant, for example, salivary glands. These OARs were considered during treatment planning for both modalities according to the dose constraints at each centre, see appendix B tables B.4 and B.5. However, they were not contoured for all patients enrolled in this study, as for the majority of brain tumour patients, the salivary glands would be exposed to zero or very low doses. If the model-based approach would be applied to this patient subgroup further NTCP models could be considered, for example, xerostomia after irradiation of the parotid gland (Beetz et al., 2012c).

6.4 Summary

The present study showed, that cranial PBT can achieve a substantial relative dose reduction compared to XRT in a variety of OARs, especially in contralateral organs. However, if the dose in contralateral structures is quite low for both modalities, the large relative dose reduction may be less clinically relevant. Thus, treatment decisions based on dosimetric plan comparisons have to consider relative and absolute dose differences as well as the absolute dose value. A model-based selection of patients for PBT is based on NTCP predictions for both modalities. For the majority of patients and a variety of side effects, NTCP predictions were higher for XRT than for PBT, but the NTCP differences did often not exceed a threshold of 10 percentage points for an individual side effect. However, based on the model selection in the presented study, a majority of patients would still have been selected for PBT if the treatment decision would be based on this threshold value for at least one side effect. The available NTCP models for side effects following cranial RT should be used with caution in clinical practice as they are subject to a variety of uncertainties and limitations. One of the most decisive side effects for the treatment modality decision may be neurocognitive dysfunction. Thus, future studies on XRT and PBT should focus on comprehensive testing of neurocognition to develop valid NTCP models for this side effect. For extensive data collection, electronic medical charts and structured databases are essential. Prior to the application of NTCP models for treatment decision, the applied thresholds should be examined intensively by conducting dose and NTCP comparing studies.

The presented feasibility study underlines that physical dose-volume parameters or metrics alone may not be sufficient to describe the clinical relevance between different treatment plans and techniques. However, physical parameters have the advantage of being determined with relatively high accuracy. To estimate the clinical impact of the dosimetric differences, NTCP models may be more appropriate. The major limitations of this approach, however, are the model uncertainties which add to the dosimetric uncertainties. In order to reduce these uncertainties, further NTCP models for different modern treatment techniques or even re-irradiated patients are needed, allowing the radiobiological approach to demonstrate its validity in clinical practice.

7 Conclusion and further perspectives

Radiotherapy is an essential component in the treatment of patients with intracranial tumours. Compared to XRT techniques, PBT offers the possibility to reduce the integral dose and the volume of healthy tissue irradiated with low and intermediate doses due to the physical properties of protons. Although the number of PBT centres is still increasing, their capacity remains limited due to the high personnel, technical and therefore financial expenditure. Thus, patients need to be identified that may benefit most from treatment with PBT. Dosimetric comparisons suggest that PBT may be beneficial in two different aspects. First, dose could be escalated in radioresistant tumours, increasing tumour control while keeping potential side effects at a level comparable to non-escalated XRT regimes, e.g. in some lesions in the skull base. Second, a reduction in long-term side effects could be expected at the same tumour coverage compared to XRT, since the healthy tissue surrounding the tumour is less affected. However, up to now, there are no data from RCTs available that demonstrate the superiority of PBT over modern XRT in brain tumour patients.

In Germany, the reimbursement of costs varies between health insurance companies. Some conduct individual case examination, while others have agreements with single PBT facilities for specific tumour entities. In other European countries, the allocation of limited PBT capacities among patients that are likely to benefit from PBT is organised differently. In the Netherlands, the nationwide model-based approach has been implemented by the National Health Care Institute. For patients that are considered to benefit most from PBT according to the model-based selection procedure, the costs of PBT will be reimbursed irrespective of the health care insurance company (Langendijk et al., 2018).

In this thesis, it was investigated whether the model-based approach, which has so far mainly been used for head and neck cancer patients, can be extended to patients with intracranial tumours. For this purpose, data from a total of 305 patients diagnosed with LGG, HGG, meningioma and other benign tumours, were evaluated. All patients were treated with PBT at one of three different PBT facilities. NTCP models for acute and late physician-rated side effects following cranial PBT have been developed and externally validated. Dose-volume parameters of the skin characterising high doses were associated with early alopecia grade ≥ 2 and erythema grade ≥ 1. Validation confirmed that the presented dose-volume parameters were strongly associated with these endpoints for all investigated cohorts, but the calibration differed. Thus, centre-specific differences in toxicity assessment should be considered carefully. Late alopecia grade ≥ 1 was also associated with high dose-volume parameters of the skin while late hearing impairment grade ≥ 1 was related to cochlear mean doses. Additionally, in analyses of pooled data from the three PBT centres, dose-response relations between late physician-rated memory impairment and the dose to the remaining brain and the hippocampi as well as late fatigue and

dose to the brain stem were observed. Due to the relatively low incidence rates and low severity grades of late side effects, NTCP models could only be developed for mild forms. In a model-based approach for PBT patient selection, however, only models that predict side effects of grade ≥ 2 would be considered (Langendijk et al., 2018). Dose-response relationships for alopecia and dry eye syndrome of grade ≥ 2 were found in GEE analyses, but they still need to be validated externally.

The preservation of neurocognitive function after RT is an important aim of treatment, as this side effect may considerably reduce patients' quality of life. Hence, the neurocognitive function up to two years after cranial PBT was investigated in this thesis. Self-reported and objectively measured neurocognitive function (MoCA test) and most QoL domains remained largely stable over time during recurrence-free follow-up after PBT. Slight deterioration of the MoCA score was associated with an increased relative volume of the irradiated anterior cerebellum exposed to doses of 30 to 40 Gy(RBE). Larger studies, such as the ongoing comparative clinical trials (NCT02824731, 2016; NCT03180502, 2017), may allow for validation of these findings, development of NTCP models and a direct comparison of neurocognitive changes between patients treated with either cranial PBT or XRT.

An integral part of the model-based patient selection for PBT is an *in silico* treatment plan comparison between a photon and proton treatment plan to estimate the potential benefit of PBT over XRT for an individual patient. In this thesis, a comparison of PBT and VMAT plans was conducted for 92 patients. Additionally, the created treatment plans and eleven published NTCP models predicting different side effects were used to investigate the feasibility of the model-based approach for patients with intracranial tumours. It was observed that PBT can achieve a substantial relative dose reduction compared to XRT in many OARs, especially in contralateral organs. However, if the absolute dose in contralateral structures is quite low for both modalities, the large relative dose reduction may be of less clinical relevance. Thus, treatment decisions based on dosimetric plan comparisons have to consider relative and absolute dose differences as well as the absolute dose value. Corresponding to these dosimetric results, NTCP predictions for the majority of patients and a variety of side effects were higher for XRT than for PBT, but often their difference did not exceed a threshold of 10 percentage points. However, since several models have been investigated, the majority of patients would finally have been selected for PBT if treatment decision was based upon a 10 percentage point threshold for at least one side effect. It is obvious that the number of selected patients for PBT largely depends on the applied thresholds for NTCP reduction. To avoid heterogeneity in clinical trials selecting patients for PBT according to the model-based approach, the thresholds have to be consistent inter-institutional (Grau, 2013).

Furthermore, patient selection highly depends on the created treatment plans for both RT techniques. Often, treatment plan quality greatly varies between institutions or planners (Das et al., 2008; Nelms et al., 2012a; Tol et al., 2014; Das et al., 2017). In the future, automated planning systems enabling standardised planning may be used for this purpose, such as Rapid-

Plan™(Varian Medical Systems, Palo Alto, CA). These knowledge-planning solutions could also improve the clinical implementation of the model-based patient selection, as two treatment plans for each patient are required (Delaney et al., 2017). Comparisons of PBT patients with a historical cohort of photon therapy patients should be treated carefully, as photon RT techniques are still evolving.

The presented feasibility study underlines that physical dose-volume parameters alone may not be sufficient to describe the clinical relevance between different treatment plans and techniques. To estimate the clinical impact of dosimetric differences, NTCP models may be more appropriate. However, physical parameters have the advantage of being determined with relatively high accuracy (depending on dose algorithm, positioning and contouring accuracy, etc.), while for NTCP predictions, comparatively large model uncertainties add to dosimetric uncertainties. Even with validated models, these uncertainties may strongly influence the accuracy of patient selection (Bijman et al., 2017). To reduce these effects, further NTCP models for different modern treatment techniques are needed, allowing the radiobiological approach to demonstrate its validity in clinical practice.

Further perspectives

The introduction of the model-based approach for the selection of patients with intracranial tumours for PBT into clinical practice requires further detailed investigations. To calibrate or develop new NTCP models, large prospective clinical trials comprising patients of different treatment modalities need to be conducted. Reliable NTCP models require additional high-quality toxicity data including extensive pre- and post-treatment testing, such as endocrine function tests or audiograms to measure hearing ability. Comprehensive standardised test batteries should be used to assess neurocognitive dysfunction. In contrast to physician-rated side effect assessment, such tests are not subject to interobserver variability and may therefore be more precise. But since they are considerably more time-consuming than a physician's evaluation, it should be focused on the most decisive side effects. To reduce the required time and manpower, larger follow-up intervals could be performed for a follow-up of several years, e.g. biannually or even yearly for extended testing procedures for neurocognitive function. In multi-centre studies, it should be ensured that all participating centres have the same follow-up intervals.

Other alternatives to physician-rated study endpoints or objective measures are QoL measures. Although these are highly subjective, they reflect the actual condition of the patient without interpretation by a physician. Detailed follow-up may also comprise specific questionnaires to address the subjective perception of neurocognitive function, e.g. the FACT-Cog questionnaire.

Prospective electronic recording of side effects (and their absence) as well as patient and treatment-specific information such as comorbidities, medication, clinical cofactors and years of education, is important to access and combine these data with imaging data and RT treatment parameters easily.

In addition to the collection of extensive outcome data, a precise determination of the dose-volume parameters is important for NTCP modelling. This also includes consistent contouring of the OARs to be examined. To reduce contouring variability between physicians, consistent use of delineation guidelines and atlases is essential. Additionally, the implementation of auto-contouring tools could be considered (Vinod et al., 2016). These tools are currently being further improved, e.g. by deep-learning-based segmentation approaches (van der Heyden et al., 2019; van Dijk et al., 2020). Besides reducing the delineation time, auto-contouring can also facilitate the segmentation of additional OARs. These could be used in NTCP modelling to investigate unknown relationships between dose to these structures and side effects such as neurocognitive function.

Machine learning approaches have been widely used for survival data analysis based on imaging data in the field of radiation oncology (Parmar et al., 2015; Leger et al., 2017). It has been proposed to extend these methods to normal tissue response by combining features describing the dose distributions (dosiomics), image features of the OARs (radiomics), genetic variations leading to alterations in radiosensitivity (genomics) and biomarkers characterising inflammatory reactions as well as clinical characteristics (Alsner et al., 2008; El Naqa et al., 2009; Lambin et al., 2013b; Gulliford, 2015; Dean et al., 2017; El Naqa et al., 2017; El Naqa et al., 2018; Gabryś et al., 2018). Several models using machine learning techniques have been developed to predict different side effects such as genitourinary and gastrointestinal toxicity (Pella et al., 2011; Ospina et al., 2014), radiation-induced pneumonitis (Chen et al., 2007; Valdes et al., 2016) or xerostomia (Buettner et al., 2012; Gabryś et al., 2018). With these approaches, different challenges of NTCP modelling, such as a large number of possible predictors, varying follow-up times, group imbalance due to low incidence rates, and heterogeneous and noisy data, could be handled (Gabryś et al., 2018). However, new challenges are arising, such as the choice of classifier and feature selection algorithms, as they may impact the predictive performance of the models. As these advanced modelling techniques are less transparent, it may be more difficult to implement them in clinical planning routines. In the future, machine learning could offer new perspectives for NTCP modelling but still requires high-quality input data as with classical NTCP modelling. However, to benefit from machine learning algorithms compared to normal logistic regression approaches, much larger patient cohorts are needed, since no advantage has been demonstrated for small patient cohorts so far (Gabryś et al., 2018).

In addition to physician-rated side-effects, further objective radiation-induced effects on MR images could be assessed during follow-up. Although MR image changes do not necessarily reflect the clinical condition of the patient, focal image changes may develop into radiation-induced brain necrosis, a severe side effect. The development of NTCP models for radiation-induced necrosis is challenging because a clear distinction between necrotic tissue and tumour progression is difficult and makes the definition of the modelling endpoint complex. As necrosis occurs as very

local single or multifocal lesions, it is difficult to define a dedicated OAR for modelling. Radiation-induced enhancing foci were observed in the periventricular region (Shah et al., 2012; Eulitz et al., 2019a), and this area may be delineated routinely in the future to develop and validate models based on this structure. Although radiation-induced necrosis is a rare side effect, reports of brain necrosis after PBT indicate possible biological differences between XRT and PBT (Haas-Kogan et al., 2018). As this endpoint can be so severe for the patient, it should be considered in a model-based approach in the future.

In clinical practice, RBE is so far considered constant, although it is known that in proton beams both LET and RBE are variable and increase at the distal edge of the spread-out Bragg peak (Paganetti, 2014; Lühr et al., 2018). For treatment plans with several beam directions, LET/RBE may superimpose and further increase in the overlap area leading to enhanced biological effects. These effects may be pronounced in intensity-modulated PBT due to highly altered LET and RBE values (Haas-Kogan et al., 2018). In future, the development of NTCP models for PBT could not only consider the physical dose distribution (weighted with the constant RBE factor of 1.1.) but also LET or RBE distributions.

It can be concluded from this thesis that the introduction of the model-based approach for the selection of patients with intracranial tumours for PBT is feasible. However, some effort is still required for its implementation. In a nutshell, the following steps should be realised in the future:

- Careful selection of clinically most relevant side effects to be considered in a model-based approach; in particular: choice of rather objectively measurable side effects, such as neurocognitive function with extensive test batteries and relevant changes on MRI (severe brain necrosis) as well as the selection of subjective QoL endpoints
- Long-term studies including these side effects as well as distinctive endpoints (and thus also tests and grading systems) of published NTCP models for calibration or validation
- Extension of the follow-up intervals (e.g. biannually) to ensure practical implementation
- Electronic recording of all patient- and treatment-relevant parameters and endpoints
- Determination of appropriate $\Delta NTCP$ threshold values for the selected NTCP models, depending on available PBT capacities

This thesis has highlighted that the extent of clinical benefits may vary greatly on a patient-individual level. By developing different NTCP models and investigating neurocognitive function following PBT, this thesis provided a basis for future studies on treatment selection enabling the best possible patient care.

8 Zusammenfassung

Hintergrund

Die Strahlentherapie ist ein wesentlicher Bestandteil der Behandlung von Patienten mit intrakraniellen Tumoren. Im Vergleich zur konventionellen Photonentherapie bietet die Protonentherapie aufgrund der physikalischen Eigenschaften von Protonen die Möglichkeit, das Volumen des gesunden Gewebes, das mit niedrigen und mittleren Dosen bestrahlt wird, zu reduzieren. Anhand dosimetrischer Vergleiche lassen sich potentielle Vorteile der Protonentherapie im Vergleich zur Photonentherapie ableiten. Beispielsweise könnte bei gleicher Tumorerfassung eine Reduktion der Langzeitnebenwirkungen erwartet werden, da das gesunde Gewebe im Vergleich zur Photonentherapie weniger stark belastet wird. Bislang liegen jedoch keine Daten aus randomisierten Studien vor, die eine Überlegenheit der Protonentherapie gegenüber modernen Photonentherapietechniken bei Hirntumorpatienten nahelegen würden. Obwohl die Zahl der Protonentherapiezentren weiter steigt, bleiben ihre Kapazitäten aufgrund des hohen technischen, personellen und damit finanziellen Aufwands begrenzt. In Deutschland hängt die Kostenerstattung für eine Protonentherapie primär von den Krankenkassen ab, während in anderen europäischen Ländern die limitierten Protonentherapieplätze landesweit zentral verteilt werden. In den Niederlanden wurde der modellbasierte Ansatz zur Auswahl der Patienten, die am meisten von einer Protonentherapie profitieren können, vorgeschlagen und umgesetzt. Dabei werden anhand von Modellen die Wahrscheinlichkeiten für Normalgewebskomplikationen (NTCP) eines Protonen- und eines Photonentherapieplans für jeden Patienten individuell abgeschätzt. Liegt die potentielle NTCP-Reduktion des Protonentherapieplans über einem definierten Schwellwert, z.B. 10 Prozentpunkte, wird der Patient anhand dieses Bestrahlungsplans mit Protonen behandelt. Allerdings existieren bisher nur wenige dieser obligatorischen NTCP-Modelle für verschiedene Nebenwirkungen nach kranialer Strahlentherapie, insbesondere für die Protonentherapie.

Fragestellung

Im Rahmen dieser Arbeit wurde untersucht, inwieweit der modellbasierte Ansatz, der bisher vor allem für Patienten mit Kopf-Hals-Tumoren verwendet wurde, auf Patienten mit intrakraniellen Tumoren ausgedehnt werden kann. Zu diesem Zweck wurden verschiedene NTCP-Modelle für Früh- und Spätnebenwirkungen entwickelt und extern validiert. Bei Hirntumorpatienten stellt außerdem der Erhalt der neurokognitiven Funktion nach einer Strahlentherapie ein wichtiges Therapieziel dar, da diese Nebenwirkung die Lebensqualität der Patienten erheblich beeinträchtigen kann. Daher wurde die neurokognitive Funktion von Patienten, die mit kranialer PBT behandelt wurden, untersucht, um mögliche Dosis-Wirkungs-Beziehungen zu erkennen. Als integraler Bestandteil des modellbasierten Ansatzes wurde ein patientenindividueller Planvergleich durchge-

führt, um den potenziellen Nutzen der Protonentherapie gegenüber der Photonentherapie abzuschätzen.

Material und Methoden

Es wurden Daten von insgesamt 305 Patienten ausgewertet, bei denen niedrig- oder hochgradige Gliome, Meningiome oder andere gutartige, intrakranielle Tumoren diagnostiziert wurden. Alle Patienten wurden in einem von drei verschiedenen Protonentherapiezentren behandelt. Diverse NTCP-Modelle für akute und späte Nebenwirkungen wurden mittels logistischer Regression anhand klinischer Nachsorgedaten sowie dosimetrischer Parameter assoziierter Risikoorgane erstellt. Diese Modelle wurden extern validiert. Die neurokognitive Funktion bis zu zwei Jahre nach kranialer Protonentherapie wurde mit Hilfe von Fragebögen zur subjektiven Lebensqualität und dem objektiven Montreal Cognitive Assessment (MoCA) Test untersucht. Veränderungen der neurokognitiven Funktion wurden mit dosimetrischen Parametern verschiedener Hirnstrukturen in gemischten Modellen analysiert. Ein Vergleich von Protonen- mit Photonentherapieplänen, die mit einer volumenmodulierten Technik erstellt wurden, wurde bei 92 Patienten durchgeführt. Zusätzlich wurde anhand dieser Behandlungspläne sowie elf veröffentlichter NTCP-Modelle für verschiedene Nebenwirkungen die Realisierbarkeit des modellbasierten Ansatzes für Patienten mit intrakraniellen Tumoren untersucht.

Ergebnisse

Frühe Alopezie Grad $\geq$ 2 und Erytheme Grad $\geq$ 1 waren mit Dosis-Volumen-Parametern der Haut assoziiert, die den Hochdosisbereich charakterisieren. Die Validierung der Modelle bestätigte, dass diese Dosis-Volumen-Parameter bei allen untersuchten Patientenkohorten stark mit diesen Endpunkten assoziiert waren. Da sich jedoch Abweichungen in der Kalibrierung zeigten, sollten zentrumsspezifische Unterschiede in der Bewertung und Vorhersage von Nebenwirkungswahrscheinlichkeiten sorgfältig berücksichtigt werden. Späte Alopezie Grad $\geq$ 1 war ebenfalls mit Hochdosisparametern der Haut assoziiert, während späte Hörminderungen Grad $\geq$ 1 mit der mittleren Dosis der ipsilateralen Cochlea in Verbindung standen. Darüber hinaus wurden bei Analyse der gepoolten Daten aller drei Protonentherapiezentren Zusammenhänge zwischen spät auftretenden Gedächtnisstörungen und der Dosis im nichtbetroffenen Gehirn sowie der Hippocampi beobachtet. Ebenso wurde ein Zusammenhang zwischen spät auftretender Fatigue und dem Hochdosisbereich im Hirnstamm gefunden, jedoch müssen diese Zusammenhänge noch anhand externer Kohorten validiert werden.

Sowohl die subjektiven Einschätzungen als auch die objektiv gemessene neurokognitive Funktion blieben innerhalb der rezidivfreien Nachbeobachtungszeit weitgehend stabil. Eine leichte Verschlechterung des MoCA-Wertes war mit einem erhöhten relativen Volumen des bestrahlten anterioren Cerebellums assoziiert, das einer Dosis von 30 bis 40 Gy(RBW) ausgesetzt war.

Der Vergleich der Bestrahlungspläne ergab, dass mit der Protonentherapie im Vergleich zur Photonentherapie in vielen, insbesondere in kontralateralen, Risikoorganen eine erhebliche relative Dosisreduktion erreichen werden konnte. Fällt die absolute Dosis in kontralateralen Strukturen jedoch bei beiden Modalitäten recht niedrig aus, ist die große relative Dosisreduktion möglicherweise von geringerer klinischer Relevanz. Behandlungsentscheidungen auf Grundlage von dosimetrischen Planvergleichen sollten deshalb sowohl relative und absolute Dosisunterschiede als auch die Beträge der absoluten Dosiswerte berücksichtigen. Die Vorhersagen der NTCP-Werte waren für die Mehrheit der Patienten und eine Reihe von Nebenwirkungen für die Photonentherapie höher als für die Protonentherapie. Für die meisten untersuchten Nebenwirkungen und Patienten lagen die NTCP-Differenzen jedoch nicht über einem Schwellwert von 10 Prozentpunkten. Trotzdem wäre anhand der Untersuchungen in dieser Arbeit eine Mehrheit der Patienten für die Protonentherapie ausgewählt worden, da viele verschiedene NTCP-Modelle untersucht wurden und bereits durch Schwellwertüberschreitung bei einem einzigen Modell eine Protonentherapie indiziert wäre.

Schlussfolgerung

Die vorgestellte Analyse unterstreicht, dass der modellbasierte Ansatz als Therapieentscheidungsstrategie zwischen Photonen- oder Protonentherapie auch für Patienten mit intrakraniellen Tumoren anwendbar sein kann, da NTCP-Vorhersagen besser geeignet sind als physikalische Dosis-Volumen-Parameter, um die klinische Relevanz zwischen verschiedenen Behandlungstechniken abzuschätzen. NTCP-Vorhersagen unterliegen jedoch im Vergleich zu physikalischen Parametern sowohl dosimetrischen als auch Modellunsicherheiten, die die Genauigkeit der Patientenauswahl stark beeinflussen können. Neurokognitive Funktionsstörungen könnten in Kombination mit anderen akuten und späten Nebenwirkungen eine der entscheidenden Nebenwirkungen sein, die bei der modellbasierten Patientenauswahl berücksichtigt werden sollten. Vor der Implementierung des Ansatzes für Patienten mit intrakraniellen Tumoren sind jedoch weitere Langzeitstudien mit objektiven und subjektiven Charakteristika notwendig, um die essentiellen NTCP-Modelle zu validieren und geeignete Schwellwerte zu definieren. Darüber hinaus sollten auch Variationen der radiobiologischen Wirksamkeit beachtet werden, z.B. um strahleninduzierte Hirnnekrosen adäquat zu berücksichtigen.

Diese Arbeit hat gezeigt, dass das Ausmaß des klinischen Nutzens der Protonentherapie patientenindividuell sehr unterschiedlich sein kann. Durch die Entwicklung verschiedener NTCP-Modelle und die Untersuchung der neurokognitiven Funktion nach Protonentherapie wurde eine Grundlage für zukünftige Entwicklungen einer personalisierten Behandlungsentscheidung geschaffen.

9 Summary

Background

Radiotherapy is an essential component in the treatment of patients with intracranial tumours. Compared to conventional photon therapy (XRT) techniques, proton beam therapy (PBT) offers the possibility to reduce the volume of healthy tissue irradiated with low and intermediate doses due to the physical properties of protons. Dosimetric comparisons suggest that PBT may be beneficial in different aspects, e.g. by reducing side effects at the same tumour coverage compared to XRT. However, up to now, there are no data from randomised controlled trials available that demonstrate the superiority of PBT over modern XRT in brain tumour patients. Although the number of PBT centres is still increasing, their capacity remains limited due to the high technical, personnel and therefore financial expenditure. In Germany, the reimbursement of costs varies between health insurance companies, while in other European countries, the allocation of limited PBT capacities is organised in a nationwide standard procedure. In the Netherlands, the model-based approach has been suggested and implemented to select patients that may benefit most from treatment with PBT. Using normal tissue complication probability (NTCP) models, the complication risks of a proton and photon therapy plan are estimated for each patient. If the potential NTCP reduction of the PBT plan exceeds a predefined threshold, e.g. 10 percentage points, the patient is treated with protons. However, mandatory NTCP models for different side effects following cranial radiotherapy are rare, especially for PBT.

Objectives

This thesis investigated whether the model-based approach, which has so far mainly been used for head and neck cancer patients, can be extended to patients with intracranial tumours. For this purpose, different NTCP models for early and late physician-rated side effects were developed and externally validated. Moreover, for brain tumour patients, the preservation of neurocognitive function after radiotherapy is an especially important aim of treatment, as this side effect may considerably reduce patients' quality of life. Hence, the neurocognitive function of patients treated with cranial PBT was investigated to reveal potential dose-effect relationships. As an integral part of the model-based patient approach, an *in silico* treatment plan comparison was conducted to estimate the presumed benefit of PBT over XRT for individual patients.

Material and methods

Data from a total of 305 patients diagnosed with low-grade and high-grade glioma, meningioma and other benign intracranial tumours were evaluated. All patients were treated with PBT at one of three different PBT institutes. NTCP models for early and late physician-rated side effects have

been created using logistic regression based on clinical outcome data and dosimetric parameters of associated organs at risk. These models were externally validated.

Neurocognitive function up to two years after cranial PBT was investigated using subjective quality of life questionnaires and the objective Montreal Cognitive Assessment (MoCA) test. Changes in the neurocognitive function were related to dosimetric parameters of different brain structures using mixed model analyses.

The treatment plan comparison was conducted in 92 patients. Dosimetric parameters of different organs at risk were compared between retrospectively created XRT plans (volumetric modulated arc therapy) and the clinically applied PBT plan. Additionally, these treatment plans and eleven published NTCP models were used to estimate potential complication risks and to investigate the feasibility of the model-based approach for patients with intracranial tumours.

Results

Dose-volume parameters of the skin characterising high doses were associated with early alopecia grade $\geq$ 2 and erythema grade $\geq$ 1. Validation confirmed that the presented dose-volume parameters were strongly associated with these endpoints for all investigated cohorts, but the calibration differed. Thus, centre-specific differences in toxicity assessment should be considered carefully. Late alopecia grade $\geq$ 1 was also associated with high dose-volume parameters of the skin, while late hearing impairment grade $\geq$ 1 was related to cochlear mean doses. Additionally, in analyses of pooled data from the three PBT centres, dose-response relations between late physician-rated memory impairment and the dose to the remaining brain and the hippocampi as well as late fatigue and dose to the brain stem were observed, but still need to be validated externally.

Self-reported and objectively measured neurocognitive function (MoCA test) and most quality of life domains remained largely stable over time during recurrence-free follow-up after PBT. Slight deterioration of the MoCA score was associated with an increased relative volume of the irradiated anterior cerebellum exposed to doses of 30 to 40 Gy(RBE).

Treatment plan comparison revealed that PBT is able to achieve a substantial relative dose reduction compared to XRT in many organs at risk, especially in contralateral organs. However, if the absolute dose in contralateral structures is quite low for both modalities, the large relative dose reduction may be of less clinical relevance. Thus, treatment decisions based on dosimetric plan comparisons have to consider relative and absolute dose differences as well as the absolute dose value. Corresponding to these dosimetric results, NTCP predictions for the majority of patients and a variety of side effects were higher for XRT than for PBT, but often their difference did not exceed a threshold of 10 percentage points. However, since several models have been investigated, the majority of patients would finally have been selected for proton beam therapy if treatment decision was based upon a 10 percentage point threshold for at least one side effect.

Conclusion

The present feasibility study underlines that the model-based approach could be a suitable treatment decision strategy between proton or photon therapy for patients with intracranial tumours, as NTCP predictions may be more appropriate than physical dose-volume parameters to estimate the clinical relevance between different treatment techniques. However, NTCP predictions are subject to both dosimetric and model uncertainties which may strongly influence the accuracy of patient selection. Neurocognitive dysfunction could be one of the decisive side effects included in the model-based patient selection, in combination with other acute and late side effects. Before implementing the approach for patients with intracranial tumours, further long-term studies using objective and subjective measures to define patient outcome are necessary to validate the mandatory NTCP models and to define appropriate thresholds for treatment decision. Furthermore, variations in radiobiological effectiveness should also be considered, e.g. to adequately address radiation-induced brain necrosis.

This thesis highlighted that the extent of the clinical benefit of PBT may vary greatly on a patient-individual level. By developing different NTCP models and investigating neurocognitive function following PBT, this thesis provided a basis for future studies on treatment selection enabling the best possible patient care.

Bibliography

Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, Haes JCJMd, Kaasa S, Klee M, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw K, Sullivan M, Takeda F. 1993. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life Instrument for use in international clinical trials in oncology. J Natl Cancer Inst 85:365–376.

Aaronson NK, Taphoorn MJB, Heimans JJ, Postma TJ, Gundy CM, Beute GN, Slotman BJ, Klein M. 2011. Compromised health-related quality of life in patients with low-grade glioma. J Clin Oncol 29:4430–4435.

Adeberg S, Harrabi SB, Bougatf N, Bernhardt D, Rieber J, Koerber SA, Syed M, Sprave T, Mohr A, Abdollahi A, Haberer T, Combs SE, Herfarth K, Debus J, Rieken S. 2016. Intensity-modulated proton therapy, volumetric-modulated arc therapy, and 3D conformal radiotherapy in anaplastic astrocytoma and glioblastoma. Strahlenther Onkol 192:770–779.

Adeberg S, Harrabi SB, Bougatf N, Verma V, Windisch P, Bernhardt D, Combs SE, Herfarth K, Debus J, Rieken S. 2018. Dosimetric comparison of proton radiation therapy, volumetric modulated arc therapy, and three-dimensional conformal radiotherapy based on intracranial tumor location. Cancers 10:401.

Adeberg S, Harrabi SB, Verma V, Bernhardt D, Grau N, Debus J, Rieken S. 2017. Treatment of meningioma and glioma with protons and carbon ions. Radiat Oncol 12:193.

Agresti A. 2002. Categorical data analysis. John Wiley & Sons, New York.

Akpati H, Kim C, Kim B, Park T, Meek A. 2008. Unified dosimetry index (UDI): A figure of merit for ranking treatment plans. J Appl Clin Med Phys 9:99–108.

Alfonso JCL, Herrero MA, Núñez L. 2015. A dose-volume histogram based decision-support system for dosimetric comparison of radiotherapy treatment plans. Radiat Oncol 10:263.

Alsner J, Andreassen CN, Overgaard J. 2008. Genetic markers for prediction of normal tissue toxicity after radiotherapy. Semin Radiat Oncol 18:126–135.

Arjomandy B, Sahoo N, Cox J, Lee A, Gillin M. 2009. Comparison of surface doses from spot scanning and passively scattered proton therapy beams. Phys Med Biol 54:N295.

Armstrong CL, Hunter JV, Ledakis GE, Cohen B, Tallent EM, Goldstein BH, Tochner Z, Lustig R, Judy KD, Pruitt A, Mollman JE, Stanczak EM, Jo MY, Than TL, Phillips P. 2002. Late cognitive and radiographic changes related to radiotherapy. Neurology 59:40.

Armstrong CL, Shera DM, Lustig RA, Phillips PC. 2012. Phase measurement of cognitive impairment specific to radiotherapy. Int J Radiat Oncol Biol Phys 83:e319–e324.

Barker JL, Jr, Garden AS, Ang KK, O'Daniel JC, Wang H, Court LE, Morrison WH, Rosenthal DI, Chao KSC, Tucker SL, Mohan R, Dong L. 2004. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. Int J Radiat Oncol Biol Phys 59:960–970.

Basch E, Reeve BB, Mitchell SA, Clauser SB, Minasian LM, Dueck AC, Mendoza TR, Hay J, Atkinson TM, Abernethy AP, Bruner DW, Cleeland CS, Sloan JA, Chilukuri R, Baumgartner P, Denicoff A, St. Germain D, O'Mara AM, Chen A, Kelaghan J, Bennett AV, Sit L, Rogak L, Barz A, Paul DB, Schrag D. 2014. Development of the national cancer institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). J Natl Cancer Inst 106:dju244.

Batth SS, Sreeraman R, Dienes E, Beckett LA, Daly ME, Cui J, Mathai M, Purdy JA, Chen AM. 2013. Clinical-dosimetric relationship between lacrimal gland dose and ocular toxicity after intensity-modulated radiotherapy for sinonasal tumours. Br J Radiol 86:20130459.

Baumann M, Grégoire V. 2009. Modified fractionation. In: Joiner MC, van der Kogel AJ (Eds) Basic clinical radiobiology. 4th ed. Hodder Education, London, pp. 41–55.

Baumert BG, Lomax AJ, Miltchev V, Davis JB. 2001. A comparison of dose distributions of proton and photon beams in stereotactic conformal radiotherapy of brain lesions. Int J Radiat Oncol Biol Phys 49:1439–1449.

Beetz I, Schilstra C, Burlage FR, Koken PW, Doornaert P, Bijl HP, Chouvalova O, Leemans CR, de Bock GH, Christianen MEMC, van der Laan BFAM, Vissink A, Steenbakkers RJHM, Langendijk JA. 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 ME, Doornaert P, Bijl HP, Chouvalova O, van den Heuvel ER, Steenbakkers RJ, Langendijk JA. 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.

Beetz I, Schilstra C, van der Schaaf A, van den Heuvel ER, Doornaert P, van Luijk P, Vissink A, van der Laan BF, Leemans CR, Bijl HP, Christianen ME, Steenbakkers RJ, Langendijk JA. 2012c. NTCP models for patient-rated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: The role of dosimetric and clinical factors. Radiother Oncol 105:101–106.

Béhin A, Delattre JY. 2003. Neurologic sequelae of radiotherapy on the nervous system. In: Schiff D, Wen P (Eds) Cancer neurology in clinical practice. Humana Press, Totowa, NJ, pp. 173–191.

Bekelman JE, Hahn SM. 2014. Reference pricing with evidence development: A way forward for proton therapys. J Clin Oncol 32:1540–1542.

Bender ET. 2012. Brain necrosis after fractionated radiation therapy: Is the halftime for repair longer than we thought? Med Phys 39:7055–7061.

van den Bent MJ, Afra D, de Witte O, Hassel MB, Schraub S, Hoang-Xuan K, Malmström PO, Collette L, Piérart M, Mirimanoff R, Karim ABMF. 2005. Long-term efficacy of early versus delayed radiotherapy for low-grade astrocytoma and oligodendroglioma in adults: The EORTC 22845 randomised trial. Lancet 366:985–990.

van den Bent MJ, Baumert B, Erridge SC, Vogelbaum MA, Nowak AK, Sanson M, Brandes AA, Clement PM, Baurain JF, Mason WP, Wheeler H, Chinot OL, Gill S, Griffin M, Brachman DG, Taal W, Rudà R, Weller M, McBain C, Reijneveld J, Enting RH, Weber DC, Lesimple T, Clenton S, Gijtenbeek A, Pascoe S, Herrlinger U, Hau P, Dhermain F, van Heuvel I, Stupp R, Aldape K, Jenkins RB, Dubbink HJ, Dinjens WNM, Wesseling P, Nuyens S, Golfinopoulos V, Gorlia T, Wick W, Kros JM. 2017a. Interim results from the CATNON trial (EORTC study 26053-22054) of treatment with concurrent and adjuvant temozolomide for 1p/19q non-co-deleted anaplastic glioma: A phase 3, randomised, open-label intergroup study. Lancet 390:1645–1653.

van den Bent MJ, Smits M, Kros JM, Chang SM. 2017b. Diffuse infiltrating oligodendroglioma and astrocytoma. J Clin Oncol 35:2394–2401.

Bentzen SM. 2008. Randomized controlled trials in health technology assessment: Overkill or overdue? Radiother Oncol 86:142–147.

Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, Ten Haken RK, Yorke ED. 2010. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): An introduction to the scientific issues. Int J Radiat Oncol Biol Phys 76:S3–S9.

Bhandare N, Antonelli PJ, Morris CG, Malayapa RS, Mendenhall WM. 2007. Ototoxicity after radiotherapy for head and neck tumors. Int J Radiat Oncol Biol Phys 67:469–479.

Bhandare N, Jackson A, Eisbruch A, Pan CC, Flickinger JC, Antonelli P, Mendenhall WM. 2010. Radiation therapy and hearing loss. Int J Radiat Oncol Biol Phys 76:S50–S57.

Bhandare N, Moiseenko V, Song WY, Morris CG, Bhatti MT, Mendenhall WM. 2012. Severe dry eye syndrome after radiotherapy for head-and-neck tumors. Int J Radiat Oncol Biol Phys 82:1501–1508.

Bijman RG, Breedveld S, Arts T, Astreinidou E, de Jong MA, Granton PV, Petit SF, Hoogeman MS. 2017. Impact of model and dose uncertainty on model-based selection of oropharyngeal cancer patients for proton therapy. Acta Oncol 56:1444–1450.

Black PM, Villavicencio AT, Rhouddou C, Loeffler JS. 2001. Aggressive surgery and focal radiation in the management of meningiomas of the skull base: Preservation of function with maintenance of local control. Acta Neurochir (Wien) 143:555–562.

Blanchard P, Wong AJ, Gunn GB, Garden AS, Mohamed AS, Rosenthal DI, Crutison J, Wu R, Zhang X, Zhu XR, Mohan R, Amin MV, Fuller CD, Frank SJ. 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.

Boele FW, Zant M, Heine ECE, Aaronson NK, Taphoorn MJB, Reijneveld JC, Postma TJ, Heimans JJ, Klein M. 2014. The association between cognitive functioning and health-related quality of life in low-grade glioma patients. Neurooncol Pract 1:40–46.

Bolsi A, Fogliata A, Cozzi L. 2003. Radiotherapy of small intracranial tumours with different advanced techniques using photon and proton beams: A treatment planning study. Radiother Oncol 68:1–14.

Bosma I, Vos MJ, Heimans JJ, Taphoorn MJB, Aaronson NK, Postma TJ, van der Ploeg HM, Muller M, Vandertop WP, Slotman BJ, Klein M. 2007. The course of neurocognitive functioning in high-grade glioma patients. Neuro Oncol 9:53–62.

Bottomley A, Flechtner H, Efficace F, Vanvoorden V, Coens C, Therasse P, Velikova G, Blazeby J, Greimel E. 2005. Health related quality of life outcomes in cancer clinical trials. Eur J Cancer 41:1697–1709.

Bower JE, Ganz PA, Irwin MR, Kwan L, Breen EC, Cole SW. 2011. Inflammation and behavioral symptoms after breast cancer treatment: Do fatigue, depression, and sleep disturbance share a common underlying mechanism? J Clin Oncol 29:3517–3522.

Brizel DM, Light K, Zhou SM, Marks LB. 1999. Conformal radiation therapy treatment planning reduces the dose to the optic structures for patients with tumors of the paranasal sinuses. Radiother Oncol 51:215–218.

Brouwer CL, Steenbakkers RJ, Gort E, Kamphuis ME, van der Laan HP, van't Veld AA, Sijtsema NM, Langendijk JA. 2014. Differences in delineation guidelines for head and neck cancer result in inconsistent reported dose and corresponding NTCP. Radiother Oncol 111:148–152.

Brown AP, Barney CL, Grosshans DR, McAleer MF, de Groot JF, Puduvalli VK, Tucker SL, Crawford CN, Khan M, Khatua S, Gilbert MR, Brown PD, Mahajan A. 2013. Proton beam cran-

iospinal irradiation reduces acute toxicity for adults with medulloblastoma. Int J Radiat Oncol Biol Phys 86:277–284.

Brown PD, Buckner JC, Uhm JH, Shaw EG. 2003. The neurocognitive effects of radiation in adult low-grade glioma patients. Neuro Oncol 5:161–167.

Brundage MD, Pater JL, Zee B. 1993. Assessing the reliability of two toxicity scales: Implications for interpreting toxicity data. J Natl Cancer Inst 85:1138–1148.

Buckner JC, Shaw EG, Pugh SL, Chakravarti A, Gilbert MR, Barger GR, Coons S, Ricci P, Bullard D, Brown PD, Stelzer K, Brachman D, Suh JH, Schultz CJ, Bahary JP, Fisher BJ, Kim H, Murtha AD, Bell EH, Won M, Mehta MP, Curran WJ. 2016. Radiation plus procarbazine, CCNU, and vincristine in low-grade glioma. N Engl J Med 374:1344–1355.

Buettner F, Miah AB, Gulliford SL, Hall E, Harrington KJ, Webb S, Partridge M, Nutting CM. 2012. Novel approaches to improve the therapeutic index of head and neck radiotherapy: An analysis of data from the PARSPORT randomised phase III trial. Radiother Oncol 103:82–87.

Burman C, Kutcher G, Emami B, Goitein M. 1991. Fitting of normal tissue tolerance data to an analytic function. Int J Radiat Oncol Biol Phys 21:123–135.

Burns PB, Rohrich RJ, Chung KC. 2011. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 128:305–310.

Casanueva FF, Molitch ME, Schlechte JA, Abs R, Bonert V, Bronstein MD, Brue T, Cappabianca P, Colao A, Fahlbusch R, Fideleff H, Hadani M, Kelly P, Kleinberg D, Laws E, Marek J, Scanlon M, Sobrinho LG, Wass JAH, Giustina A. 2006. Guidelines of the Pituitary Society for the diagnosis and management of prolactinomas. Clin Endocrinol (Oxf) 65:265–273.

Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, Silberman M, Yellen SB, Winicour P, Brannon J. 1993. The Functional Assessment of Cancer Therapy scale: Development and validation of the general measure. J Clin Oncol 11:570–579.

Cella L, Palma G, Deasy JO, Oh JH, Liuzzi R, D'Avino V, Conson M, Pugliese N, Picardi M, Salvatore M, Pacelli R. 2014. Complication probability models for radiation-induced heart valvular dysfunction: Do heart-lung interactions play a role? PLoS One 9:1–11.

Chaichana KL, Jusue-Torres I, Navarro-Ramirez R, Raza SM, Pascual-Gallego M, Ibrahim A, Hernandez-Hermann M, Gomez L, Ye X, Weingart JD, Olivi A, Blakeley J, Gallia GL, Lim M, Brem H, Quinones-Hinojosa A. 2014. Establishing percent resection and residual volume thresholds affecting survival and recurrence for patients with newly diagnosed intracranial glioblastoma. Neuro Oncol 16:113–122.

Chan E, Khan S, Oliver R, Gill SK, Werring DJ, Cipolotti L. 2014. Underestimation of cognitive impairments by the Montreal Cognitive Assessment (MoCA) in an acute stroke unit population. J Neurol Sci 343:176–179.

Chen S, Zhou S, Yin FF, Marks LB, Das SK. 2007. Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis. Med Phys 34:3808–3814.

Chen WC, Jackson A, Budnick AS, Pfister DG, Kraus DH, Hunt MA, Stambuk H, Levegrun S, Wolden SL. 2006. Sensorineural hearing loss in combined modality treatment of nasopharyngeal carcinoma. Cancer 106:820–829.

Chera BS, Amdur RJ, Patel P, Mendenhall WM. 2009. A radiation oncologists guide to contouring the hippocampus. Am J Clin Oncol 32:20–22.

Christianen ME, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, Doornaert P, Koken PW, Leemans CR, Rinkel RN, de Bruijn MJ, de Bock G, Roodenburg JL, van der Laan BF, Slotman BJ, Verdonck-de Leeuw IM, Bijl HP, Langendijk JA. 2012. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: Results of a prospective observational study. Radiother Oncol 105:107–114.

Cohen J. 1960. A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46.

Cole AM, Scherwath A, Ernst G, Lanfermann H, Bremer M, Steinmann D. 2013. Self-reported cognitive outcomes in patients with brain metastases before and after radiation therapy. Int J Radiat Oncol Biol Phys 87:705–712.

Combs SE. 2017. Does proton therapy have a future in CNS tumors? Curr Treat Options Neurol 19:12.

Connor M, Karunamuni R, McDonald C, White N, Pettersson N, Moiseenko V, Seibert T, Marshall D, Cervino L, Bartsch H, Kuperman J, Murzin V, Krishnan A, Farid N, Dale A, Hattangadi-Gluth J. 2016. Dose-dependent white matter damage after brain radiotherapy. Radiother Oncol 121:209–216.

Constine LS, Woolf PD, Cann D, Mick G, McCormick K, Raubertas RF, Rubin P. 1993. Hypothalamic-pituitary dysfunction after radiation for brain tumors. N Engl J Med 328:87–94.

Cooley SA, Heaps JM, Bolzenius JD, Salminen LE, Baker LM, Scott SE, Paul RH. 2015. Longitudinal change in performance on the montreal cognitive assessment in older adults. Clin Neuropsychol 29:824–835.

Correa DD. 2010. Neurocognitive function in brain tumors. Curr Neurol Neurosci Rep 10:232–239.

Correa DD, DeAngelis LM, Shi W, Thaler HT, Lin M, Abrey LE. 2007. Cognitive functions in low-grade gliomas: Disease and treatment effects. J Neurooncol 81:175–184.

Danesh-Meyer HV. 2010. Radiation-induced optic neuropathy. J Clin Neurosci 15:95–100.

Darzy KH, Shalet SM. 2009. Hypopituitarism following radiotherapy. Pituitary 12:40–50.

Das IJ, Andersen A, Chen Z, Dimofte A, Glatstein E, Hoisak J, Huang L, Langer MP, Lee C, Pacella M, Popple RA, Rice R, Smilowitz J, Sponseller P, Zhu T. 2017. State of dose prescription and compliance to international standard (ICRU-83) in intensity modulated radiation therapy among academic institutions. Pract Radiat Oncol 7:e145–e155.

Das IJ, Cheng CW, Chopra KL, Mitra RK, Srivastava SP, Glatstein E. 2008. Intensity-modulated radiation therapy dose prescription, recording, and delivery: Patterns of variability among institutions and treatment planning systems. J Natl Cancer Inst 100:300–307.

Dawson LA, Biersack M, Lockwood G, Eisbruch A, Lawrence TS, Ten Haken RK. 2005. Use of principal component analysis to evaluate the partial organ tolerance of normal tissues to radiation. Int J Radiat Oncol Biol Phys 62:829–837.

De Marzi L, Feuvret L, Boulé T, Habrand J, Martin F, Calugaru V, Fournier-Bidoz N, Ferrand R, Mazal A. 2015. Use of gEUD for predicting ear and pituitary gland damage following proton and photon radiation therapy. Br J Radiol 88:20140413.

Dean JA, Welsh LC, Wong KH, Aleksic A, Dunne E, Islam MR, Patel A, Patel P, Petkar I, Phillips I, Sham J, Schick U, Newbold KL, Bhide SA, Harrington KJ, Nutting CM, Gulliford SL. 2017. Normal tissue complication probability (NTCP) modelling of severe acute mucositis using a novel oral mucosal surface organ at risk. Clin Oncol 29:263–273.

Delaney AR, Dahele M, Tol JP, Kuijper IT, Slotman BJ, Verbakel WFAR. 2017. Using a knowledge-based planning solution to select patients for proton therapy. Radiother Oncol 124:263–270.

Dennis ER, Bussière MR, Niemierko A, Lu MW, Fullerton BC, Loeffler JS, Shih HA. 2013. A comparison of critical structure dose and toxicity risks in patients with low grade gliomas treated with IMRT versus proton radiation therapy. Technol Cancer Res T 12:1–9.

Di Maio M, Gallo C, Leighl NB, Piccirillo MC, Daniele G, Nuzzo F, Gridelli C, Gebbia V, Ciardiello F, De Placido S, Ceribelli A, Favaretto AG, de Matteis A, Feld R, Butts C, Bryce J, Signoriello S, Morabito A, Rocco G, Perrone F. 2015. Symptomatic toxicities experienced during anticancer treatment: Agreement between patient and physician reporting in three randomized trials. J Clin Oncol 33:910–915.

Diallo I, Haddy N, Adjadj E, Samand A, Quiniou E, Chavaudra J, Alziar I, Perret N, Guérin S, Lefkopoulos D, de Vathaire F. 2009. Frequency distribution of second solid cancer locations in relation to the irradiated volume among 115 patients treated for childhood cancer. Int J Radiat Oncol Biol Phys 74:876–883.

van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, Steenbakkers RJ, Langendijk JA, Gooding MJ, Brouwer CL. 2020. Improving automatic delineation for head and neck organs at risk by deep learning contouring. Radiother Oncol 142:115–123.

Douw L, Klein M, Fagel SSAA, van den Heuvel J, Taphoorn MJB, Aaronson NK, Postma TJ, Vandertop WP, Mooij JJ, Boerman RH, Beute GN, Sluimer JD, Slotman BJ, Reijneveld JC, Heimans JJ. 2009. Cognitive and radiological effects of radiotherapy in patients with low-grade glioma: Long-term follow-up. Lancet Neurol 8:810–818.

DRKS00004384. 2012. Register study standard proton therapy WPE- Adults - (ProReg). German Clinical Trials Center [Internet], Cologne: German Institute of Medical Documentation and Information (Germany), 2008. Identifier: DRKS00004384 [registered 2012 Sep 28, cited 2019 Jul 10]. URL: https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00004384.

DRKS00007670. 2015. Proton therapy of brain tumors: Prospective collection of effectiveness and side effects with standard clinical doses (Proto-R-Hirn). German Clinical Trials Center [Internet], Cologne: German Institute of Medical Documentation and Information (Germany), 2008. Identifier DRKS00007670 [registered 2015 Feb 02, cited 2019 Jul 10]. URL: https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00007670.

DRKS00008569. 2015. Proton therapy of skull base tumors: Prospective collection of effectiveness and side effects with standard clinical doses (Proto-R-Schädelbasis). German Clinical Trials Center [Internet], Cologne: German Institute of Medical Documentation and Information (Germany), 2008. Identifier DRKS00008569 [registered 2015 May 21, cited 2019 Jul 10]. URL: https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00008569.

Durante M, Orecchia R, Loeffler JS. 2017. Charged-particle therapy in cancer: Clinical uses and future perspectives. Nat Rev Clin Oncol 14:483–495.

Dutz A, Agolli L, Troost EGC, Krause M, Baumann M, Lühr A, Löck S. 2017a. EP-1595: NTCP models for early toxicities in patients with prostate or brain tumours receiving proton therapy. Radiother Oncol 123:S860.

Dutz A, Agolli L, Troost E, Krause M, Baumann M, Lühr A, Vermeren X, Geismar D, Timmermann B, Löck S. 2018. OC-0511: NTCP modelling and external validation of early side effects for proton therapy of brain tumours. Radiother Oncol 127:S266–S267.

Dutz A, Agolli L, Bütof R, Valentini C, Baumann M, Lühr A, Löck S, Krause M. 2020. Neurocognitive function and quality of life after proton beam therapy for brain tumour patients. Radiother Oncol. [Epub ahead of print] DOI: 10.1016/j.radonc.2019.12.024.

Dutz A, Agolli L, Troost E, Krause M, Baumann M, Lühr A, Löck S. 2017b. Modelling of NTCP for acute side effects in patients with prostate cancer or brain tumours receiving proton therapy. Biomed Tech (Berl) 62:s304.

Dutz A, Lühr A, Agolli L, Troost EGC, Krause M, Baumann M, Vermeren X, Geismar D, Schapira EF, Bussière M, Daly JE, Bussière MR, Timmermann B, Shih HA, Löck S. 2019. Development and validation of NTCP models for acute side-effects resulting from proton beam therapy of brain tumours. Radiother Oncol 130:164–171.

Eekers DBP, in 't Ven L, Deprez S, Jacobi L, Roelofs E, Hoeben A, Lambin P, de Ruysscher D, Troost EGC. 2017. The posterior cerebellum, a new organ at risk? Clin Transl Radiat Oncol 8:22–26.

Eekers DBP, in 't Ven L, Roelofs E, Postma A, Alapetite C, Burnet NG, Calugaru V, Compter I, Coremans IEM, Høyer M, Lambrecht M, Nyström PW, Romero AM, Paulsen F, Perpar A, de Ruysscher D, Renard L, Timmermann B, Vitek P, Weber DC, van der Weide HL, Whitfield GA, Wiggenraad R, Troost EGC. 2018. The EPTN consensus-based atlas for CT- and MR-based contouring in neuro-oncology. Radiother Oncol 128:37–43.

Eekers DBP, Roelofs E, Jelen U, Kirk M, Granzier M, Ammazzalorso F, Ahn PH, Janssens GORJ, Hoebers FJP, Friedmann T, Solberg T, Walsh S, Troost EGC, Kaanders JHAM, Lambin P. 2016. Benefit of particle therapy in re-irradiation of head and neck patients. Results of a multicentric *in silico* ROCOCO trial. Radiother Oncol 121:387–394.

El Naqa I, Bradley JD, Lindsay PE, Hope AJ, Deasy JO. 2009. Predicting radiotherapy outcomes using statistical learning techniques. Phys Med Biol 54:S9–S30.

El Naqa I, Johansson A, Owen D, Cuneo K, Cao Y, Matuszak M, Bazzi L, Lawrence TS, Ten Haken RK. 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.

El Naqa I, Kerns SL, Coates J, Luo Y, Speers C, West CML, Rosenstein BS, Ten Haken RK. 2017. Radiogenomics and radiotherapy response modeling. Phys Med Biol 62:R179–R206.

Emami B, Lyman J, Brown A, Cola L, Goitein M, Munzenrider JE, Shank B, Solin LJ, Wesson M. 1991. Tolerance of normal tissue to therapeutic irradiation. Int J Radiat Oncol Biol Phys 21:109–122.

Eulitz J, Troost E, Raschke F, Schulz E, Lutz B, Dutz A, Löck S, Wohlfahrt P, Enghardt W, Karpowitz C, Krause M, Lühr A. 2019a. Predicting late magnetic resonance image changes in glioma patients after proton therapy. Acta Oncol 58:1536–1539.

Eulitz J, Lutz B, Wohlfahrt P, Dutz A, Enghardt W, Karpowitz C, Krause M, Troost EGC, Lühr A. 2019b. A Monte Carlo based radiation response modelling framework to assess variability of clinical RBE in proton therapy. Phys Med Biol 64:225020.

Fairchild A, Langendijk JA, Nuyts S, Scrase C, Tomsej M, Schuring D, Gulyban A, Ghosh S, Weber DC, Budach W. 2014. Quality assurance for the EORTC 22071-26071 study: Dummy run prospective analysis. Radiat Oncol 9:248.

Fayers P, Aaronson N, Bjordal K, Groenvold M, Curran D, Bottomley A. 2001. The EORTC QLQ-C30 scoring manual (3rd Edition). European Organisation for Research and Treatment of Cancer, Brussels.

Ferris MJ, Zhong J, Switchenko JM, Higgins KA, Cassidy RJ, McDonald MW, Eaton BR, Patel KR, Steuer CE, Baddour HM, Miller AH, Bruner DW, Xiao C, Beitler JJ. 2018. Brainstem dose is associated with patient-reported acute fatigue in head and neck cancer radiation therapy. Radiother Oncol 126:100–106.

Ferrufino-Ponce ZK, Henderson BA. 2006. Radiotherapy and cataract formation. Semin Ophthalmol 21:171–180.

Feuvret L, Noël G, Mazeron JJ, Bey P. 2006. Conformity index: A review. Int J Radiat Oncol Biol Phys 64:333–342.

Fitzek MM, Thornton AF, Harsh G, Rabinov JD, Munzenrider JE, Lev M, Ancukiewicz M, Bussiere M, Hedley-Whyte ET, Hochberg FH, Pardo FS. 2001. Dose-escalation with proton/photon irradiation for Daumas-Duport lower-grade glioma: Results of an institutional phase I/II trial. Int J Radiat Oncol Biol Phys 51:131–137.

Folstein MF, Folstein SE, McHugh PR. 1975. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12:189–198.

Foppiano F, Fiorino C, Frezza G, Greco C, Valdagni R. 2003. The impact of contouring uncertainty on rectal 3D dose–volume data: Results of a dummy run in a multicenter trial (AIRO-PROS01–02). Int J Radiat Oncol Biol Phys 57:573–579.

Fossati P, Vavassori A, Deantonio L, Ferrara E, Krengli M, Orecchia R. 2016. Review of photon and proton radiotherapy for skull base tumours. Rep Pract Oncol Radiother 21:336–355.

Frost JA, Binder JR, Springer JA, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW. 1999. Language processing is strongly left lateralized in both sexes: Evidence from functional MRI. Brain 122:199–208.

Gabryś HS, Buettner F, Sterzing F, Hauswald H, Bangert M. 2018. Design and selection of machine learning methods using radiomics and dosiomics for normal tissue complication probability modeling of xerostomia. Front Oncol 8:35.

Gagnon G, Hansen K, Woolmore-Goodwin S, Gutmanis I, Wells J, Borrie M, Fogarty J. 2013. Correcting the MoCA for education: Effect on sensitivity. Can J Neurol Sci 40:678–683.

Gamer M, Lemon J, Singh I. 2010. irr: Various Coefficients of Interrater Reliability and Agreement.

Ginot A, Doyen J, Hannoun-Lévi JM, Courdi A. 2010. Dose de tolérance des tissus sains : La peau et les phanères. Cancer Radiother 14:379–385.

Glaudemans AWJM, Enting RH, Heesters MAAM, Dierckx RAJO, van Rheenen RWJ, Walenkamp AME, Slart RHJA. 2013. Value of ^{11}C-methionine PET in imaging brain tumours and metastases. Eur J Nucl Med Mol Imaging 40:615–635.

Glimelius B, Ask A, Bjelkengren G, Björk-Eriksson T, Blomquist E, Johansson B, Karlsson M, Zackrisson B. 2009. Number of patients potentially eligible for proton therapy. Acta Oncol 44:836–849.

Glimelius B, Montelius A. 2007. Proton beam therapy – Do we need the randomised trials and can we do them? Radiother Oncol 83:105–109.

Goitein M, Cox JD. 2008. Should randomized clinical trials be required for proton radiotherapy? J Clin Oncol 26:175–176.

Gondi V, Hermann BP, Mehta MP, Tomé WA. 2012. Hippocampal dosimetry predicts neurocognitive function impairment after fractionated stereotactic radiotherapy for benign or low-grade adult brain tumors. Int J Radiat Oncol Biol Phys 83:e487–e493.

Gondi V, Pugh SL, Tome WA, Caine C, Corn B, Kanner A, Rowley H, Kundapur V, DeNittis A, Greenspoon JN, Konski AA, Bauman GS, Shah S, Shi W, Wendland M, Kachnic L, Mehta MP. 2014. Preservation of memory with conformal avoidance of the hippocampal neural stem-cell compartment during whole-brain radiotherapy for brain metastases (RTOG 0933): A phase II multi-institutional trial. J Clin Oncol 32:3810–3816.

Gordon KB, Char DH, Sagerman RH. 1995. Late effects of radiation on the eye and ocular adnexa. Int J Radiat Oncol Biol Phys 31:1123–1139.

Goyal R, Blood AJ, Potters L, Kapur A. 2015. The effect of assessment criteria on inter-rater variability in the evaluation of skin reactions following breast cancer radiation therapy. In: Jaffray

DA (Ed) World congress on medical physics and biomedical engineering, June 7-12, 2015, Toronto, Canada. IFMBE Proceedings. Springer, Cham, pp. 501–504.

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.

Greenberger BA, Pulsifer MB, Ebb DH, MacDonald SM, Jones RM, Butler WE, Huang MS, Marcus KJ, Oberg JA, Tarbell NJ, Yock TI. 2014. Clinical outcomes and late endocrine, neurocognitive, and visual profiles of proton radiation for pediatric low-grade gliomas. Int J Radiat Oncol Biol Phys 89:1060–1068.

Grimaldi G, Manto M. 2012. Topography of cerebellar deficits in humans. Cerebellum 11:336–351.

Grosshans DR, Mohan R, Gondi V, Shih HA, Mahajan A, Brown PD. 2017. The role of image-guided intensity modulated proton therapy in glioma. Neuro Oncol 19:ii30–ii37.

Grosshans DR, Zhu XR, Melancon A, Allen PK, Poenisch F, Palmer M, McAleer MF, McGovern SL, Gillin M, DeMonte F, Chang EL, Brown PD, Mahajan A. 2014. Spot scanning proton therapy for malignancies of the base of skull: treatment planning, acute toxicities, and preliminary clinical outcomes. Int J Radiat Oncol Biol Phys 90:540–546.

Grubbs JRJ, Tolleson-Rinehart S, Huynh K, Davis RM. 2014. A review of quality of life measures in dry eye questionnaires. Cornea 33:215–218.

Gulliford S. 2015. Modelling of normal tissue complication probabilities (NTCP): Review of application of machine learning in predicting NTCP. In: El Naqa I, Li R, Murphy MJ (Eds) Machine learning in radiation oncology: Theory and applications. Springer, Cham, pp. 277–310.

Gulliford SL, Miah AB, Brennan S, McQuaid D, Clark CH, Partridge M, Harrington KJ, Morden JP, Hall E, Nutting CM. 2012a. Dosimetric explanations of fatigue in head and neck radiotherapy: An analysis from the PARSPORT Phase III trial. Radiother Oncol 104:205–212.

Gulliford SL, Partridge M, Sydes MR, Webb S, Evans PM, Dearnaley DP. 2012b. Parameters for the Lyman Kutcher Burman (LKB) model of Normal Tissue Complication Probability (NTCP) for specific rectal complications observed in clinical practise. Radiother Oncol 102:347–351.

Haas-Kogan D, Indelicato D, Paganetti H, Esiashvili N, Mahajan A, Yock T, Flampouri S, MacDonald S, Fouladi M, Stephen K, Kalapurakal J, Terezakis S, Kooy H, Grosshans D, Makrigiorgos M, Mishra K, Poussaint TY, Cohen K, Fitzgerald T, Gondi V, Liu A, Michalski J, Mirkovic D, Mohan R, Perkins S, Wong K, Vikram B, Buchsbaum J, Kun L. 2018. National cancer institute workshop on proton therapy for children: Considerations regarding brainstem injury. Int J Radiat Oncol Biol Phys 101:152–168.

de Haes JC, van Knippenberg FC, Neijt JP. 1990. Measuring psychological and physical distress in cancer patients: structure and application of the Rotterdam Symptom Checklist. Br J Cancer 62:1034–1038.

Harrabi SB, Bougatf N, Mohr A, Haberer T, Herfarth K, Combs SE, Debus J, Adeberg S. 2016. Dosimetric advantages of proton therapy over conventional radiotherapy with photons in young patients and adults with low-grade glioma. Strahlenther Onkol 192:759–769.

Hauswald H, Rieken S, Ecker S, Kessel KA, Herfarth K, Debus J, Combs SE. 2012. First experiences in treatment of low-grade glioma grade I and II with proton therapy. Radiat Oncol 7:189.

Henk J, Whitelocke R, Warrington A, Bessell E. 1993. Radiation dose to the lens and cataract formation. Int J Radiat Oncol Biol Phys 25:815–820.

van der Heyden B, Wohlfahrt P, Eekers DBP, Richter C, Terhaag K, Troost EGC, Verhaegen F. 2019. Dual-energy CT for automatic organs-at-risk segmentation in brain-tumor patients using a multi-atlas and deep-learning approach. Sci Rep 9:4126.

Hickok JT, Morrow GR, McDonald S, Bellg AJ. 1996. Frequency and correlates of fatigue in lung cancer patients receiving radiation therapy: Implications for management. J Pain Symptom Manage 11:370–377.

Ho WK, Wei WI, Kwong DL, Sham JS, Tai PT, Yuen AP, Au DK. 1999. Long-term sensorineural hearing deficit following radiotherapy in patients suffering from nasopharyngeal carcinoma: A prospective study. Head Neck 21:547–553.

Höller U, Tribius S, Kuhlmey A, Grader K, Fehlauer F, Alberti W. 2003. Increasing the rate of late toxicity by changing the score? A comparison of RTOG/EORTC and LENT/SOMA scores. Int J Radiat Oncol Biol Phys 55:1013–1018.

Holthusen H. 1936. Erfahrungen über die Verträglichkeitsgrenze für Röntgenstrahlen und deren Nutzanwendung zur Verhütung von Schäden. Strahlentherapie 57:254–269.

Honoré HB, Bentzen SM, Møller K, Grau C. 2002. Sensori-neural hearing loss after radiotherapy for nasopharyngeal carcinoma: Individualized risk estimation. Radiother Oncol 65:9–16.

Hug EB, DeVries A, Thornton AF, Munzenrider JE, Pardo FS, Hedley-Whyte ET, Bussiere MR, Ojemann R. 2000. Management of atypical and malignant meningiomas: Role of high-dose, 3D-conformal radiation therapy. J Neurooncol 48:151–160.

ICRU. 1999. ICRU Report 62 - Prescribing, recording and reporting photon beam therapy (Supplement to ICRU Report 50). Bethesda, MD, USA.

ICRU. 2007. ICRU Report 78 - Prescribing, recording and reporting proton-beam therapy. J ICRU 7:1–210.

ICRU. 2010. ICRU Report 83 - Prescribing, recording, and reporting photon-beam intensity modulated radiation therapy (IMRT). J ICRU 10:1–106.

Jacobs M, Boersma L, Dekker A, Swart R, Lambin P, de Ruysscher D, Verhaegen F, Stultiens J, Ramaekers B, van Merode F. 2017. What is the impact of innovation on output in healthcare with a special focus on treatment innovations in radiotherapy? A literature review. Br J Radiol 90:20170251.

Jakobi A. 2015. Evaluation of proton treatment strategies for head and neck cancer and lung cancer based on treatment planning studies. PhD thesis. OncoRay – National Center for Radiation Research in Oncology, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden.

Jaspers J, Mèndez Romero A, Hoogeman MS, van den Bent M, Wiggenraad RGJ, Taphoorn MJB, Eekers DBP, Lagerwaard FJ, Lucas Calduch AM, Baumert BG, Klein M. 2019. Evaluation of the hippocampal normal tissue complication model in a prospective cohort of low grade glioma patients-an analysis within the EORTC 22033 clinical trial. Front Oncol 9:991.

Jeganathan VSE, Wirth A, MacManus MP. 2011. Ocular risks from orbital and periorbital radiation therapy: A critical review. Int J Radiat Oncol Biol Phys 79:650–659.

Jereczek-Fossa BA, Zarowski A, Milani F, Orecchia R. 2003. Radiotherapy-induced ear toxicity. Cancer Treat Rev 29:417–430.

Jereczek-Fossa BA, Marsiglia HR, Orecchia R. 2002. Radiotherapy-related fatigue. Crit Rev Oncol Hematol 41:317–325.

Jhaveri J, Cheng E, Tian S, Buchwald Z, Chowdhary M, Liu Y, Gillespie TW, Olson JJ, Diaz AZ, Voloschin A, Eaton BR, Crocker IR, McDonald MW, Curran WJ, Patel KR. 2018. Proton vs. photon radiation therapy for primary gliomas: An analysis of the National Cancer Data Base. Front Oncol 8:440.

Joiner M, van der Kogel A, Steel G. 2009. Introduction: the significance of radiobiology and radiotherapy for cancer treatment. In: Joiner MC, van der Kogel AJ (Eds) Basic clinical radiobiology. 4th ed. Hodder Education, London, pp. 41–55.

Joiner MC. 2009. Quantifying cell kill and cell survival. In: Joiner MC, van der Kogel AJ (Eds) Basic clinical radiobiology. 4th ed. Hodder Education, London, pp. 41–55.

Joiner MC, Bentzen SM. 2009. Fractionation: The linear-quadratic approach. In: Joiner MC, van der Kogel AJ (Eds) Basic clinical radiobiology. 4th ed. Hodder Education, London, pp. 11–26.

Julayanont P, Phillips N, Chertkow H, Nasreddine ZS. 2013. Montreal Cognitive Assessment (MoCA): Concept and clinical review. In: Larner AJ (Ed) Cognitive screening instruments: A practical approach. Springer, London, pp. 111–151.

Kahalley LS, Ris MD, Grosshans DR, Okcu MF, Paulino AC, Chintagumpala M, Moore BD, Guffey D, Minard CG, Stancel HH, Mahajan A. 2016. Comparing intelligence quotient change after treatment with proton versus photon radiation therapy for pediatric brain tumors. J Clin Oncol 34:1043–1049.

Källman P, Ågren A, Brahme A. 1992. Tumour and normal tissue responses to fractionated non-uniform dose delivery. Int J Radiat Biol 62:249–262.

Kapur A, Evans C, Bloom B, Ames J, Morgenstern C, Raince J, Potters L. 2013. Interrater variability in the assessment of skin reactions in breast cancer radiation therapy: Impact of grading scales. Int J Radiat Oncol Biol Phys 87:S494.

Karger CP. 2018. Klinische Strahlenbiologie. In: Schlegel W, Karger CP, Jäkel O (Eds) Medizinische Physik. Springer Spektrum Berlin, Heidelberg, pp. 451–472.

Karim ABMF, Maat B, Hatlevoll R, Menten J, Rutten EHJM, Thomas DGT, Mascarenhas F, Horiot JC, Parvinen LM, van Reijn M, Jager JJ, Fabrini MG, van Alphen AM, Hamers HP, Gaspar L, Noordman E, Pierart M, van Glabbeke M. 1996. A randomized trial on dose-response in radiation therapy of low-grade cerebral glioma: European Organization for Research and Treatment of Cancer (EORTC) study 22844. Int J Radiat Oncol Biol Phys 36:549–556.

Kazda T, Jancalek R, Pospisil P, Sevela O, Prochazka T, Vrzal M, Burkon P, Slavik M, Hynkova L, Slampa P, Laack NN. 2014. Why and how to spare the hippocampus during brain radiotherapy: The developing role of hippocampal avoidance in cranial radiotherapy. Radiat Oncol 9:139.

van Kessel E, Baumfalk AE, van Zandvoort MJE, Robe PA, Snijders TJ. 2017. Tumor-related neurocognitive dysfunction in patients with diffuse glioma: A systematic review of neurocognitive functioning prior to anti-tumor treatment. J Neurooncol 134:9–18.

Kessel KA, Bohn C, Engelmann U, Oetzel D, Bougatf N, Bendl R, Debus J, Combs SE. 2014. Five-year experience with setup and implementation of an integrated database system for clinical documentation and research. Comput Methods Programs Biomed 114:206–217.

Kierkels RGJ, Korevaar EW, Steenbakkers RJHM, Janssen T, van't Veld AA, Langendijk JA, Schilstra C, van der Schaaf A. 2014. Direct use of multivariable normal tissue complication probability models in treatment plan optimisation for individualised head and neck cancer radiotherapy produces clinically acceptable treatment plans. Radiother Oncol 112:430–436.

Kim TG, Lim DH. 2013. Interfractional variation of radiation target and adaptive radiotherapy for totally resected glioblastoma. J Korean Med Sci 28:1233–1237.

Kleiman N. 2012. Radiation cataract. Ann ICRP 41:80–97.

Klein EE, Maserang B, Wood R, Mansur D. 2006. Peripheral doses from pediatric IMRT. Med Phys 33:2525–2531.

Klein M, Heimans JJ, Aaronson NK, van der Ploeg HM, Grit J, Muller M, Postma TJ, Mooij JJ, Boerman RH, Beute GN, Ossenkoppele GJ, van Imhoff GW, Dekker AW, Jolles J, Slotman BJ, Struikmans H, Taphoorn MJB. 2002. Effect of radiotherapy and other treatment-related factors on mid-term to long-term cognitive sequelae in low-grade gliomas: A comparative study. Lancet 360:1361–1368.

Klein M, Engelberts NHJ, van der Ploeg HM, Kasteleijn-Nolst Trenité DGA, Aaronson NK, Taphoorn MJB, Baaijen H, Vandertop WP, Muller M, Postma TJ, Heimans JJ. 2003. Epilepsy in low-grade gliomas: The impact on cognitive function and quality of life. Ann Neurol 54:514–520.

Knöös T, Wieslander E, Cozzi L, Brink C, Fogliata A, Albers D, Nyström H, Lassen S. 2006. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations. Phys Med Biol 51:5785–5807.

Kong C, Zhu XZ, Lee TF, Feng PB, Xu JH, Qian PD, Zhang LF, He X, Huang SF, Zhang YQ. 2016. LASSO-based NTCP model for radiation-induced temporal lobe injury developing after intensity-modulated radiotherapy of nasopharyngeal carcinoma. Sci Rep 6:26378.

Kong L, Lu JJ, Liss AL, Hu C, Guo X, Wu Y, Zhang Y. 2011. Radiation-induced cranial nerve palsy: A cross-sectional study of nasopharyngeal cancer patients after definitive radiotherapy. Int J Radiat Oncol Biol Phys 79:1421–1427.

Koo TK, Li MY. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15:155–163.

Krieger H. 2012. Grundlagen der Strahlungsphysik und des Strahlenschutzes. 4th ed. Vieweg + Teubner Verlag, Wiesbaden.

Krieger T, Sauer OA. 2005. Monte Carlo- versus pencil-beam-/collapsed-cone-dose calculation in a heterogeneous multi-layer phantom. Phys Med Biol 50:859–868.

Krippendorff K. 2004. Content analysis, an introduction to its methodology. 2nd ed. Thousand Oaks, CA: Sage.

Kunimatsu A, Kunimatsu N. 2017. Skull base tumors and tumor-like lesions: A pictorial review. Pol J Radiol 82:398–409.

Kutcher G, 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.

Kutcher GJ, 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.

Kwong DL, Wei WI, Sham JS, Ho W, Yuen P, Chua DT, Au DK, Wu P, Choy DT. 1996. Sensorineural hearing loss in patients treated for nasopharyngeal carcinoma: A prospective study of the effect of radiation and cisplatin treatment. Int J Radiat Oncol Biol Phys 36:281–289.

Laack NN, Brown PD. 2004. Cognitive sequelae of brain radiation in adults. Semin Oncol 31:702–713.

Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijsen J, Zegers CML, Carvalho S, Leijenaar RTH, Nalbantov G, Oberije C, Scott Marshall M, Hoebers F, Troost EGC, van Stiphout RGPM, van Elmpt W, van der Weijden T, Boersma L, Valentini V, Dekker A. 2013a. 'Rapid Learning health care in oncology' - An approach towards decision support systems enabling customised radiotherapy'. Radiother Oncol 109:159–164.

Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, Roelofs E, van Elmpt W, Boutros PC, Granone P, Valentini V, Begg AC, De Ruysscher D, Dekker A. 2013b. Predicting outcomes in radiation oncology–multifactorial decision support systems. Nat Rev Clin Oncol 10:27–40.

Lambin P, Zindler J, Vanneste B, van de Voorde L, Jacobs M, Eekers D, Peerlings J, Reymen B, Larue RTHM, Deist TM, de Jong EEC, Even AJG, Berlanga AJ, Roelofs E, Cheng Q, Carvalho S, Leijenaar RTH, Zegers CML, van Limbergen E, Berbee M, van Elmpt W, Oberije C, Houben R, Dekker A, Boersma L, Verhaegen F, Bosmans G, Hoebers F, Smits K, Walsh S. 2015. Modern clinical research: How rapid learning health care and cohort multiple randomised clinical trials complement traditional evidence based medicine. Acta Oncol 54:1289–1300.

Landier W. 2016. Ototoxicity and cancer therapy. Cancer 122:1647–1658.

Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159–174.

Langendijk JA, Boersma LJ, Rasch CRN, van Vulpen M, Reitsma JB, van der Schaaf A, Schuit E. 2018. Clinical trial strategies to compare protons with photons. Semin Radiat Oncol 28:79–87.

Langendijk JA, 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.

Lapointe S, Perry A, Butowski NA. 2018. Primary brain tumours in adults. Lancet 392:432–446.

Lawenda BD, Gagne HM, Gierga DP, Niemierko A, Wong WM, Tarbell NJ, Chen GT, Hochberg FH, Loeffler JS. 2004. Permanent alopecia after cranial irradiation: Dose-response relationship. Int J Radiat Oncol Biol Phys 60:879–887.

Lawrence YR, Li XA, el Naqa I, Hahn CA, Marks LB, Merchant TE, Dicker AP. 2010. Radiation dose-volume effects in the brain. Int J Radiat Oncol Biol Phys 76:S20–S27.

Lee TF, Fang FM, Chao PJ, Su T, Wang LK, Leung SW. 2008. Dosimetric comparisons of helical tomotherapy and step-and-shoot intensity-modulated radiotherapy in nasopharyngeal carcinoma. Radiother Oncol 89:89–96.

Lee TF, Yeh SA, Chao PJ, Chang L, Chiu CL, Ting HM, Wang HY, Huang YJ. 2015. Normal tissue complication probability modeling for cochlea constraints to avoid causing tinnitus after head-and-neck intensity-modulated radiation therapy. Radiat Oncol 10:194.

Leger S, Zwanenburg A, Pilz K, Lohaus F, Linge A, Zöphel K, Kotzerke J, Schreiber A, Tinhofer I, Budach V, Sak A, Stuschke M, Balermpas P, Rödel C, Ganswindt U, Belka C, Pigorsch S, Combs SE, Mönnich D, Zips D, Krause M, Baumann M, Troost EGC, Löck S, Richter C. 2017. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. Sci Rep 7:13206.

Li J, Bentzen SM, Li J, Renschler M, Mehta MP. 2008. Relationship between neurocognitive function and quality of life after whole-brain radiotherapy in patients with brain metastasis. Int J Radiat Oncol Biol Phys 71:64–70.

Liang KY, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73:13–22.

Liao Z, Lee JJ, Komaki R, Gomez DR, O'Reilly MS, Fossella FV, Blumenschein GR, Heymach JV, Vaporciyan AA, Swisher SG, Allen PK, Choi NC, DeLaney TF, Hahn SM, Cox JD, Lu CS, Mohan R. 2018. Bayesian adaptive randomization trial of passive scattering proton therapy and intensity-modulated photon radiotherapy for locally advanced non–small-cell lung cancer. J Clin Oncol 36:1813–1822.

Lievens Y, den Bogaert WV. 2005. Proton beam therapy: Too expensive to become true? Radiother Oncol 75:131–133.

Lievens Y, Pijls-Johannesma M. 2013. Health economic controversy and cost-effectiveness of proton therapy. Semin Radiat Oncol 23:134–141.

Light RJ. 1971. Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychol Bull 76:365–377.

Lin YS, Jen YM, Lin JC. 2002. Radiation-related cranial nerve palsy in patients with nasopharyngeal carcinoma. Cancer 95:404–409.

Liu W, Zhang X, Li Y, Mohan R. 2012. Robust optimization of intensity modulated proton therapy. Med Phys 39:1079–1091.

Loeffler JS, Shih HA. 2011. Radiation therapy in the management of pituitary adenomas. J Clin Endocrinol Metab 96:1992–2003.

van Loon EPM, Heijenbrok-Kal MH, van Loon WS, van den Bent MJ, Vincent AJPE, de Koning I, Ribbers GM. 2015. Assessment methods and prevalence of cognitive dysfunction in patients with low-grade glioma: A systematic review. J Rehabil Med 47 6:481–488.

Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW. 2016. The 2016 World Health Organization Classification of tumors of the central nervous system: A summary. Acta Neuropathol 131:803–820.

Lühr A, von Neubeck C, Krause M, Troost EGC. 2018. Relative biological effectiveness in proton beam therapy – Current knowledge and future challenges. Clin Transl Radiat Oncol 9:35–41.

Lyman JT. 1985. Complication probability as assessed from dose-volume histograms. Radiat Res Suppl 8:S13–S19.

Macbeth FR, Williams MV. 2008. Proton therapy should be tested in randomized trials. J Clin Oncol 26:2590–2591.

Macdonald DR, Kiebert G, Prados M, Yung A, Olson J. 2005. Benefit of temozolomide compared to procarbazine in treatment of glioblastoma multiforme at first relapse: Effect on neurological functioning, performance status, and health related quality of life. Cancer Invest 23:138–144.

Mahadevan A, Sampson C, LaRosa S, Floyd SR, Wong ET, Uhlmann EJ, Sengupta S, Kasper EM. 2015. Dosimetric analysis of the alopecia preventing effect of hippocampus sparing whole brain radiation therapy. Radiat Oncol 10:245.

Makale MT, McDonald CR, Hattangadi-Gluth JA, Kesari S. 2017. Mechanisms of radiotherapy-associated cognitive disability in patients with brain tumours. Nat Rev Neurol 13:52–64.

Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, Bentzen SM, Nam J, Deasy JO. 2010. The use of normal tissue complication probability (NTCP) models in the clinic. Int J Radiat Oncol Biol Phys 76:S10–S19.

Martel MK, Sandler HM, Cornblath WT, Marsh LH, Hazuka MB, Roa WH, Fraass BA, Lichter AS. 1997. Dose-volume complication analysis for visual pathway structures of patients with advanced paranasal sinus tumors. Int J Radiat Oncol Biol Phys 38:273–284.

Mayo C, Martel MK, Marks LB, Flickinger J, Nam J, Kirkpatrick J. 2010a. Radiation dose-volume effects of optic nerves and chiasm. Int J Radiat Oncol Biol Phys 76:S28–S35.

Mayo C, Yorke E, Merchant TE. 2010b. Radiation associated brainstem injury. Int J Radiat Oncol Biol Phys 76:S36–S41.

Mazzoni A, Krengli M. 2016. Historical development of the treatment of skull base tumours. Rep Pract Oncol Radiother 21:319–324.

McCorkle R, Young K. 1978. Development of a symptom distress scale. Cancer Nurs 1:373–378.

McCulloch MM, Muenz DG, Schipper MJ, Velec M, Dawson LA, Brock KK. 2018. A simulation study to assess the potential impact of developing normal tissue complication probability models with accumulated dose. Adv Radiat Oncol 3:662–672.

McDonald MW, Plankenhorn DA, McMullen KP, Henderson MA, Dropcho EJ, Shah MV, Cohen-Gadol AA. 2015. Proton therapy for atypical meningiomas. J Neurooncol 123:123–128.

Meeks SL, Buatti JM, Foote KD, Friedman WA, Bova FJ. 2000. Calculation of cranial nerve complication probability for acoustic neuroma radiosurgery. Int J Radiat Oncol Biol Phys 47:597–602.

Mendelsohn FA, Divino CM, Reis ED, Kerstein MD. 2002. Wound care after radiation therapy. Adv Skin Wound Care 15:216–224.

Menhel J, Levin D, Alezra D, Symon Z, Pfeffer R. 2006. Assessing the quality of conformal treatment planning: A new tool for quantitative comparison. Phys Med Biol 51:5363–5375.

Merchant TE, Sharma S, Xiong X, Wu S, Conklin H. 2014. Effect of cerebellum radiation dosimetry on cognitive outcomes in children with infratentorial ependymoma. Int J Radiat Oncol Biol Phys 90:547–553.

Meyers CA, Brown PD. 2006. Role and relevance of neurocognitive assessment in clinical trials of patients with CNS tumors. J Clin Oncol 24:1305–1309.

Miaskowski C. 2004. Gender differences in pain, fatigue, and depression in patients with cancer. J Natl Cancer Inst Monographs 2004:139–143.

Michelson H, Bolund C, Nilsson B, Brandberg Y. 2000. Health-related quality of life measured by the EORTC QLQ-C30: Reference values from a large sample of the Swedish population. Acta Oncol 39:477–484.

Miften MM, Das SK, Su M, Marks LB. 2004. A dose-volume-based tool for evaluating and ranking IMRT treatment plans. J Appl Clin Med Phys 5:1–14.

Milani SA, Marsiske M, Cottler LB, Chen X, Striley CW. 2018. Optimal cutoffs for the Montreal Cognitive Assessment vary by race and ethnicity. Alzheimers Dement (Amst) 10:773–781.

Molitch ME. 2017. Diagnosis and treatment of pituitary adenomas: A review. JAMA 317:516–524.

Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. 2015. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and elaboration. Ann Intern Med 162:W1–W73.

Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. 2005. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. J Am Geriatr Soc 53:695–699.

National Cancer Institute. 2009. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. National Cancer Institute, Bethesda (MD) [cited 2019 Aug 16]. URL: https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03/Archive/CTCAE_4.0_2009-05-29_QuickReference_8.5x11.pdf.

National Cancer Institute. 2017. Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0. National Cancer Institute, Bethesda (MD) [cited 2019 Aug 16]. URL: https://ctep.cancer.gov/protocolDevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf.

NCT01854554. 2013. Glioblastoma multiforme (GBM) proton vs. intensity modulated radiotherapy (IMRT). ClinicalTrials.gov [Internet], Bethesda (MD): National Library of Medicine (US), 2000 Feb 29. Identifier NCT01854554 [registered 2013 May 15, updated 2019 May 10, cited 2019 Sep 16]. URL: https://clinicaltrials.gov/ct2/show/NCT01854554.

NCT02179086. 2014. Dose-escalated photon IMRT or proton beam radiation therapy versus standard-dose radiation therapy and temozolomide in treating patients with newly diagnosed glioblastoma. ClinicalTrials.gov [Internet], Bethesda (MD): National Library of Medicine (US), 2000 Feb 29. Identifier NCT02179086 [registered 2014 July 14, updated 2019 Sep 04, cited 2019 Sep 17]. URL: https://clinicaltrials.gov/ct2/show/NCT02179086.

NCT02824731. 2016. Comparison of proton and photon radiotherapy of brain tumors (ProtoChoice-Hirn). ClinicalTrials.gov [Internet], Bethesda (MD): National Library of Medicine (US), 2000 Feb 29. Identifier NCT02824731 [registered 2016 July 07, updated 2019 Feb 18, cited 2019 June 26]. URL: https://clinicaltrials.gov/ct2/show/NCT02824731.

NCT03180502. 2017. Proton beam or intensity-modulated radiation therapy in preserving brain function in patients with IDH mutant grade II or III glioma. ClinicalTrials.gov [Internet], Bethesda (MD): National Library of Medicine (US), 2000 Feb 29. Identifier NCT03180502 [registered

2017 June 08, updated 2019 Sep 04, cited 2019 Sep 18]. URL: https://clinicaltrials.gov/ct2/show/NCT03180502.

Neben-Wittich MA, Atherton PJ, Schwartz DJ, Sloan JA, Griffin PC, Deming RL, Anders JC, Loprinzi CL, Burger KN, Martenson JA, Miller RC. 2011. Comparison of provider-assessed and patient-reported outcome measures of acute skin toxicity during a Phase III trial of mometasone cream versus placebo during breast radiotherapy: The North Central Cancer Treatment Group (N06C4). Int J Radiat Oncol Biol Phys 81:397–402.

Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, Wheeler J, Sobczak ML. 2012a. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. Pract Radiat Oncol 2:296–305.

Nelms BE, Tomé WA, Robinson G, Wheeler J. 2012b. Variations in the contouring of organs at risk: Test case from a patient with oropharyngeal cancer. Int J Radiat Oncol Biol Phys 82:368–378.

Nieder C, Langendijk JA. 2016. Normal Tissue Tolerance to Reirradiation. In: Nieder C, Langendijk J (Eds) Re-Irradiation: New Frontiers. Springer, Cham, pp. 1–15.

Niemierko A. 2006. Biological optimization. In: Bortfeld T, Schmidt-Ullrich R, De Neve W, Wazer D (Eds) Image-guided IMRT. Springer, Berlin, Heidelberg, pp. 199–216.

Niemierko A, Goitein M. 1993. Modeling of normal tissue response to radiation: The critical volume model. Int J Radiat Oncol Biol Phys 25:135–145.

Noël G, Bollet MA, Calugaru V, Feuvret L, Meder CH, Dhermain F, Ferrand R, Boisserie G, Beaudré A, Mazeron JJ, Habrand JL. 2005. Functional outcome of patients with benign meningioma treated by 3D conformal irradiation with a combination of photons and protons. Int J Radiat Oncol Biol Phys 62:1412–1422.

Noll KR, Bradshaw ME, Weinberg JS, Wefel JS. 2017. Relationships between neurocognitive functioning, mood, and quality of life in patients with temporal lobe glioma. Psychooncology 26:617–624.

Noorbakhsh A, Tang JA, Marcus LP, McCutcheon B, Gonda DD, Schallhorn CS, Talamini MA, Chang DC, Carter BS, Chen CC. 2014. Gross-total resection outcomes in an elderly population with glioblastoma: A SEER-based analysis. J Neurosurg 120:31–39.

Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, Miles EA, Miah AB, Newbold K, Tanay M, Adab F, Jefferies SJ, Scrase C, Yap BK, A'Hern RP, Sydenham MA, Emson M, Hall E. 2011. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): A phase 3 multicentre randomised controlled trial. Lancet Oncol 12:127–136.

Olson RA, Iverson GL, Carolan H, Parkinson M, Brooks BL, McKenzie M. 2011a. Prospective comparison of two cognitive screening tests: Diagnostic accuracy and correlation with community integration and quality of life. J Neurooncol 105:337.

Olson RA, Chhanabhai T, McKenzie M. 2008. Feasibility study of the Montreal Cognitive Assessment (MoCA) in patients with brain metastases. Support Care Cancer 16:1273–1278.

Olson R, Tyldesley S, Carolan H, Parkinson M, Chhanabhai T, McKenzie M. 2011b. Prospective comparison of the prognostic utility of the Mini Mental State Examination and the Montreal Cognitive Assessment in patients with brain metastases. Support Care Cancer 19:1849–1855.

Osoba D, Aaronson NK, Muller M, Sneeuw K, Hsu MA, Yung WKA, Brada M, Newlands E. 1996. The development and psychometric validation of a brain cancer quality-of-life questionnaire for use in combination with general cancer-specific questionnaires. Qual Life Res 5:139–150.

Osoba D, Rodrigues G, Myles J, Zee B, Pater J. 1998. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 16:139–144.

Ospina JD, Zhu J, Chira C, Bossi A, Delobel JB, Beckendorf V, Dubray B, Lagrange JL, Correa JC, Simon A, Acosta O, de Crevoisier R. 2014. Random forests to predict rectal toxicity following prostate cancer radiation therapy. Int J Radiat Oncol Biol Phys 89:1024–1031.

Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, Barnholtz-Sloan JS. 2017. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. Neuro Oncol 19:1–88.

Paddick I. 2000. A simple scoring ratio to index the conformity of radiosurgical treatment plans. J Neurosurg 93:219–222.

Paganetti H. 2012. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Phys Med Biol 57:R99–R117.

Paganetti H. 2014. Relative biological effectiveness (RBE) values for proton beam therapy. Variations as a function of biological endpoint, dose, and linear energy transfer. Phys Med Biol 59:R419–R472.

Paganetti H, van Luijk P. 2013. Biological considerations when comparing proton therapy with photon therapy. Semin Radiat Oncol 23:77–87.

Paganetti H, Niemierko A, Ancukiewicz M, Gerweck LE, Goitein M, Loeffler JS, Suit HD. 2002. Relative biological effectiveness (RBE) values for proton beam therapy. Int J Radiat Oncol Biol Phys 53:407–421.

Pai HH, Thornton A, Katznelson L, Finkelstein DM, Adams JA, Fullerton BC, Loeffler JS, Leibsch NJ, Klibanski A, Munzenrider JE. 2001. Hypothalamic/pituitary function following high-dose

conformal radiotherapy to the base of skull: Demonstration of a dose-effect relationship using dose-volume histogram analysis. Int J Radiat Oncol Biol Phys 49:1079–1092.

Palm RF, Oliver DE, Yang GQ, Abuodeh Y, Naghavi AO, Johnstone PAS. 2019. The role of dose escalation and proton therapy in perioperative or definitive treatment of chondrosarcoma and chordoma: An analysis of the National Cancer Data Base. Cancer 125:642–651.

Palma G, Taffelli A, Fellin F, D'Avino V, Scartoni D, Tommasino F, Scifoni E, Durante M, Amichetti M, Schwarz M, Amelio D, Cella L. 2020. Modelling the risk of radiation induced alopecia in brain tumor patients treated with scanned proton beams. Radiother Oncol 144:127–134.

Pan CC, Eisbruch A, Lee JS, Snorrason RM, Ten Haken RK, Kileny PR. 2005. Prospective study of inner ear radiation dose and hearing loss in head-and-neck cancer patients. Int J Radiat Oncol Biol Phys 61:1393–1402.

Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. 2015. Machine learning methods for quantitative radiomic biomarkers. Sci Rep 5:13087.

Parsons JT, Bova FJ, Fitzgerald CR, Mendenhall WM, Million RR. 1994. Severe dry-eye syndrome following external beam irradiation. Int J Radiat Oncol Biol Phys 30:775–780.

Particle Therapy Co-Operative Group. 2019. Particle therapy facilities in operation. [cited 2019 Sep 17]. URL: https://www.ptcog.ch/index.php/facilities-in-operation.

Pater J, Zee B, Palmer M, Johnston D, Osoba D. 1997. Fatigue in patients with cancer: Results with National Cancer Institute of Canada Clinical Trials Group studies employing the EORTC QLQ-C30. Support Care Cancer 5:410–413.

Pavy J, Denekamp J, Letschert J, Littbrand B, Mornex F, Bernier J, Gonzales-Gonzales D, Horiot J, Bolla M, Bartelink H. 1995. Late effects toxicity scoring: The SOMA scale. Int J Radiat Oncol Biol Phys 31:1043–1047.

Pedersen J, Flampouri S, Bryant C, Liang X, Mendenhall N, Li Z, Liu M, Muren LP. 2020. Cross-modality applicability of rectal normal tissue complication probability models from photon- to proton-based radiotherapy. Radiother Oncol 142:253–260.

Peeters STH, Hoogeman MS, Heemsbergen WD, Hart AAM, Koper PCM, Lebesque JV. 2006. Rectal bleeding, fecal incontinence, and high stool frequency after conformal radiotherapy for prostate cancer: Normal tissue complication probability modeling. Int J Radiat Oncol Biol Phys 66:11–19.

Pehlivan B, Ares C, Lomax AJ, Stadelmann O, Goitein G, Timmermann B, Schneider RA, Hug EB. 2012. Temporal lobe toxicity analysis after proton radiation therapy for skull base tumors. Int J Radiat Oncol Biol Phys 83:1432–1440.

Pella A, Cambria R, Riboldi M, Jereczek-Fossa BA, Fodor C, Zerini D, Torshabi AE, Cattani F, Garibaldi C, Pedroli G, Baroni G, Orecchia R. 2011. Use of machine learning methods for prediction of acute toxicity in organs at risk following prostate radiotherapy. Med Phys 38:2859–2867.

Perry JR, Laperriere N, O'Callaghan CJ, Brandes AA, Menten J, Phillips C, Fay M, Nishikawa R, Cairncross JG, Roa W, Osoba D, Rossiter JP, Sahgal A, Hirte H, Laigle-Donadey F, Franceschi E, Chinot O, Golfinopoulos V, Fariselli L, Wick A, Feuvret L, Back M, Tills M, Winch C, Baumert BG, Wick W, Ding K, Mason WP. 2017. Short-course radiation plus temozolomide in elderly patients with glioblastoma. N Engl J Med 376:1027–1037.

Peters KB, West MJ, Hornsby WE, Waner E, Coan AD, McSherry F, Herndon JE, Friedman HS, Desjardins A, Jones LW. 2014. Impact of health-related quality of life and fatigue on survival of recurrent high-grade glioma patients. J Neurooncol 120:499–506.

Piroth MD, Pinkawa M, Holy R, Stoffels G, Demirel C, Attieh C, Kaiser HJ, Langen KJ, Eble MJ. 2009. Integrated-boost IMRT or 3-D-CRT using FET-PET based auto-contoured target volume delineation for glioblastoma multiforme - a dosimetric comparison. Radiat Oncol 4:57.

Pollock BE, Stafford SL, Utter A, Giannini C, Schreiner SA. 2003. Stereotactic radiosurgery provides equivalent tumor control to Simpson Grade 1 resection for patients with small- to medium-size meningiomas. Int J Radiat Oncol Biol Phys 55:1000–1005.

Prasad G, Haas-Kogan DA. 2009. Radiation-induced gliomas. Expert Rev Neurother 9:1511–1517.

Pulsifer MB, Duncanson H, Grieco J, Evans C, Tseretopoulos ID, MacDonald S, Tarbell NJ, Yock TI. 2018. Cognitive and adaptive outcomes after proton radiation for pediatric patients with brain tumors. Int J Radiat Oncol Biol Phys 102:391–398.

van der Putten L, de Bree R, Plukker JT, Langendijk JA, Smits C, Burlage FR, Leemans CR. 2006. Permanent unilateral hearing loss after radiotherapy for parotid gland tumors. Head Neck 28:902–908.

Qi XS, Semenenko VA, Li XA. 2009. Improved critical structure sparing with biologically based IMRT optimization. Med Phys 36:1790–1799.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Raschke F, Wesemann T, Wahl H, Appold S, Krause M, Linn J, Troost E. 2019. Reduced diffusion in normal appearing white matter of glioma patients following radio(chemo)therapy. Radiother Oncol 140:110–115.

Raverot G, Jouanneau E, Trouillas J. 2014. Management of endocrine disease: Clinicopathological classification and molecular markers of pituitary tumours for personalized therapeutic strategies. Eur J Endocrinol 170:R121–R132.

Raysi Dehcordi S, Mariano M, Mazza M, Galzio R. 2013. Cognitive deficits in patients with low and high grade gliomas. J Neurosurg Sci 57:259–66.

Renovanz M, Reitzug L, Messing L, Scheurich A, Grüninger S, Ringel F, Coburger J. 2018. Patient reported feasibility and acceptance of Montreal Cognitive Assessment (MoCA) screening pre- and postoperatively in brain tumour patients. J Clin Neurosci 53:79–84.

Robinson GA, Biggs V, Walker DG. 2015. Cognitive screening in brain tumors: Short but sensitive enough? Front Oncol 5:60.

Rogers L, Barani I, Chamberlain M, Kaley TJ, McDermott M, Raizer J, Schiff D, Weber DC, Wen PY, Vogelbaum MA. 2015. Meningiomas: knowledge base, treatment outcomes, and uncertainties. A RANO review. J Neurosurg 122:4–23.

Rosewall T, Bayley AJ, Chung P, Le LW, Xie J, Baxi S, Catton CN, Currie G, Wheat J, Milosevic M. 2011. The effect of delineation method and observer variability on bladder dose-volume histograms for prostate intensity modulated radiotherapy. Radiother Oncol 101:479–485.

Rossetti HC, Lacritz LH, Cullum CM, Weiner MF. 2011. Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. Neurology 77:1272.

Roth P, Weller M. 2015. Chemotherapie von Hirntumoren bei Erwachsenen. Nervenarzt 86:495–508.

Rubin P, Constine LS, Fajardo LF, Phillips TL, Wasserman TH. 1995. Overview: Late effects of normal tissues (LENT) scoring system. Int J Radiat Oncol Biol Phys 31:1041–1042.

Salvo N, Barnes E, van Draanen J, Stacey E, Mitera G, Breen D, Giotis A, Czarnota G, Pang J, De Angelis C. 2010. Prophylaxis and management of acute radiation-induced skin reactions: A systematic review of the literature. Curr Oncol 17:94–112.

Samur AA, Coskunfirat N, Saka O. 2014. Comparison of predictor approaches for longitudinal binary outcomes: Application to anesthesiology data. PeerJ 2:e648–e648.

Schagen SB, Klein M, Reijneveld JC, Brain E, Deprez S, Joly F, Scherwath A, Schrauwen W, Wefel JS. 2014. Monitoring and optimising cognitive function in cancer patients: Present knowledge and future directions. EJC Suppl 12:29–40.

Scherman J, Appelt AL, Yu J, Persson GF, Nygård L, Langendijk JA, Bentzen SM, Vogelius IR. 2019. Incorporating NTCP into randomized trials of proton versus photon therapy. Int J Part Ther 5:24–32.

Schiavolin S, Raggi A, Scaratti C, Leonardi M, Cusin A, Visintini S, Acerbi F, Schiariti M, Zattra C, Broggi M, Ferroli P. 2018. Patients' reported outcome measures and clinical scales in brain tumor surgery: results from a prospective cohort study. Acta Neurochir (Wien) 160:1053–1061.

Schmahmann JD. 2010. The role of the cerebellum in cognition and emotion: Personal reflections since 1982 on the dysmetria of thought hypothesis, and its historical evolution from theory to therapy. Neuropsychol Rev 20:236–260.

Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, Kabani N, Toga A, Evans A, Petrides M. 1999. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. Neuroimage 10:233–260.

Schneider U. 2011. Modeling the risk of secondary malignancies after radiotherapy. Genes (Basel) 2:1033–1049.

Schwarz R, Hinz A. 2001. Reference data for the quality of life questionnaire EORTC QLQ-C30 in the general German population. Eur J Cancer 37:1345–1351.

Scoccianti S, Detti B, Gadda D, Greto D, Furfaro I, Meacci F, Simontacchi G, Di Brina L, Bonomo P, Giacomelli I, Meattini I, Mangoni M, Cappelli S, Cassani S, Talamonti C, Bordi L, Livi L. 2015. Organs at risk in the brain and their dose-constraints in adults and in children: A radiation oncologist's guide for delineation in everyday practice. Radiother Oncol 114:230–238.

Scott N, Fayers P, Aaronson N, Bottomley A, de Graeff A, Groenvold M, Gundy C, Koller M, Petersen M, Sprangers M, EORTC Quality of Life Group. 2008. EORTC QLQ-C30 Reference values manual. 2nd ed. EORTC Quality of Life Group.

Shah R, Vattoth S, Jacob R, Manzil FFP, O'Malley JP, Borghei P, Patel BN, Curé JK. 2012. Radiation necrosis in the brain: Imaging features and differentiation from tumor recurrence. Radiographics 32:1343–1359.

Shaw E, Arusell R, Scheithauer B, O'Fallon J, O'Neill B, Dinapoli R, Nelson D, Earle J, Jones C, Cascino T, Nichols D, Ivnik R, Hellman R, Curran W, Abrams R. 2002. Prospective randomized trial of low- versus high-dose radiation therapy in adults with supratentorial low-grade glioma: Initial report of a North Central Cancer Treatment Group/Radiation Therapy Oncology Group/Eastern Cooperative Oncology Group study. J Clin Oncol 20:2267–2276.

Shelley LEA, Scaife JE, Romanchikova M, Harrison K, Forman JR, Bates AM, Noble DJ, Jena R, Parker MA, Sutcliffe MPF, Thomas SJ, Burnet NG. 2017. Delivered dose can be a better predictor of rectal toxicity than planned dose in prostate radiotherapy. Radiother Oncol 123:466–471.

Shen J, Liu W, Anand A, Stoker JB, Ding X, Fatyga M, Herman MG, Bues M. 2015. Impact of range shifter material on proton pencil beam spot characteristics. Med Phys 42:1335–1340.

Sherman JC, Colvin MK, Mancuso SM, Batchelor TT, Oh KS, Loeffler JS, Yeap BY, Shih HA. 2016. Neurocognitive effects of proton radiation therapy in adults with low-grade glioma. J Neurooncol 126:157–164.

Shih HA, Loeffler JS, Tarbell NJ. 2009. Late effects of CNS radiation therapy. In: Goldman S, Turner CD (Eds) Late effects of treatment for brain tumors. Springer, Boston, MA, pp. 23–41.

Shih HA, Sherman JC, Nachtigall LB, Colvin MK, Fullerton BC, Daartz J, Winrich BK, Batchelor TT, Thornton LT, Mancuso SM, Saums MK, Oh KS, Curry WT, Loeffler JS, Yeap BY. 2015. Proton therapy for low-grade gliomas: Results from a prospective trial. Cancer 121:1712–1719.

Shrout PE, Fleiss JL. 1979. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 86 2:420–428.

Sibbald B, Roland M. 1998. Understanding controlled trials: Why are randomised controlled trials important? BMJ 316:201.

Smith AR. 2006. Proton therapy. Phys Med Biol 51:R491–R504.

Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. 2009. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. BMJ 338.

Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. 2010. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology 21:128–138.

Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, Ludwin SK, Allgeier A, Fisher B, Belanger K, Hau P, Brandes AA, Gijtenbeek J, Marosi C, Vecht CJ, Mokhtari K, Wesseling P, Villa S, Eisenhauer E, Gorlia T, Weller M, Lacombe D, Cairncross JG, Mirimanoff RO. 2009. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol 10:459–466.

Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO. 2005. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 352:987–996.

Suit H, Kooy H, Trofimov A, Farr J, Munzenrider J, DeLaney T, Loeffler J, Clasie B, Safai S, Paganetti H. 2008. Should positive phase III clinical trial data be required before proton beam therapy is more widely adopted? No. Radiother Oncol 86:148–153.

Sun Y, Yu XL, Luo W, Lee AWM, Wee JTS, Lee N, Zhou GQ, Tang LL, Tao CJ, Guo R, Mao YP, Zhang R, Guo Y, Ma J. 2014. Recommendation for a contouring method and atlas of organs at risk in nasopharyngeal carcinoma patients receiving intensity-modulated radiotherapy. Radiother Oncol 110:390–397.

Sung K, Ting H, Chao P, Guo S, Tran C, Huang Y, Wu H, Lee T. 2016. Predicting grade 2 acute radiation-induced dermatitis after hybrid intensity modulation radiotherapy for breast cancer using a logistic regression normal tissue complication probability model. Eur J Cancer 60:e4.

Surma-aho O, Niemelä M, Vilkki J, Kouri M, Brander A, Salonen O, Paetau A, Kallio M, Pyykkönen LicPhil J, Jääskeläinen J. 2001. Adverse long-term effects of brain radiotherapy in adult low-grade glioma patients. Neurology 56:1285.

Taphoorn MJB, Claassens L, Aaronson NK, Coens C, Mauer M, Osoba D, Stupp R, Mirimanoff RO, van den Bent MJ, Bottomley A. 2010. An international validation study of the EORTC brain cancer module (EORTC QLQ-BN20) for assessing health-related quality of life and symptoms in brain cancer patients. Eur J Cancer 46:1033–1040.

Taphoorn MJB, Klein M. 2004. Cognitive deficits in adult patients with brain tumours. Lancet Neurol 3:159–168.

Taylor PA, Kry SF, Alvarez P, Keith T, Lujano C, Hernandez N, Followill DS. 2016. Results from the imaging and radiation oncology core Houston's anthropomorphic phantoms used for proton therapy clinical trial credentialing. Int J Radiat Oncol Biol Phys 95:242–248.

Thakkar JP, Slevin JT, Smith CD, Sudhakar P, St Clair W, Villano JL. 2017. Bilateral radiation optic neuropathy following concurrent chemotherapy and radiation in glioblastoma. Neuroophthalmology 41:287–290.

Thomann AE, Goettel N, Monsch RJ, Berres M, Jahn T, Steiner LA, Monsch AU. 2018. The Montreal Cognitive Assessment: Normative data from a German-speaking cohort and comparison with international normative samples. J Alzheimers Dis 64:643–655.

Tol JP, Dahele M, Doornaert P, Slotman BJ, Verbakel WFAR. 2014. Different treatment planning protocols can lead to large differences in organ at risk sparing. Radiother Oncol 113:267–271.

Troeller A, Yan D, Marina O, Schulze D, Alber M, Parodi K, Belka C, Söhn M. 2015. Comparison and limitations of DVH-based NTCP models derived from 3D-CRT and IMRT data for prediction of gastrointestinal toxicities in prostate cancer patients by using propensity score matched pair analysis. Int J Radiat Oncol Biol Phys 91:435–443.

Trotti A, Bentzen SM. 2004. The need for adverse effects reporting standards in oncology clinical trials. J Clin Oncol 22:19–22.

Tsai PF, Yang CC, Chuang CC, Huang TY, Wu YM, Pai PC, Tseng CK, Wu TH, Shen YL, Lin SY. 2015. Hippocampal dosimetry correlates with the change in neurocognitive function after hippocampal sparing during whole brain radiotherapy: A prospective study. Radiat Oncol 10:253.

Valdes G, Solberg TD, Heskel M, Ungar L, Simone CB. 2016. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. Phys Med Biol 61:6105–6120.

Van den Bosch L, van der Schaaf A, van der Laan HP, Steenbakkers R, Both S, Shuit E, Wijers O, Hoebers F, Langendijk JA. 2019. Comprehensive ntcp-profiling to predict radiation-induced side effects in head and neck cancer patients treated with definitive radiotherapy. Int J Radiat Oncol Biol Phys 105:E384–E385.

Verma N, Cowperthwaite MC, Burnett MG, Markey MK. 2013. Differentiating tumor recurrence from treatment necrosis: A review of neuro-oncologic imaging strategies. Neuro Oncol 15:515–534.

Vinod S, Min M, Jameson M, Holloway L. 2016. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol 60:393–406.

Viselner G, Farina L, Lucev F, Turpini E, Lungarotti L, Bacila A, Iannalfi A, D'Ippolito E, Vischioni B, Ronchi S, Marchioni E, Valvo F, Bastianello S, Preda L. 2019. Brain MR findings in patients treated with particle therapy for skull base tumors. Insights Imaging 10:94.

Wagner L, Sweet J, Butt Z, Lai J, Cella D. 2009. Measuring patient self-reported cognitive function: Development of the functional assessment of cancer therapy–cognitive function instrument. J Support Oncol 7:W32–W39.

Wang M. 2014. Generalized estimating equations in longitudinal data analysis: A review and recent developments. Adv Stat 2014:1–11.

Weber DC, Badiyan S, Malyapa R, Albertini F, Bolsi A, Lomax AJ, Schneider R. 2016. Long-term outcomes and prognostic factors of skull-base chondrosarcoma patients treated with pencil-beam scanning proton therapy at the Paul Scherrer Institute. Neuro Oncol 18:236–243.

Weber DC, Schneider R, Goitein G, Koch T, Ares C, Geismar JH, Schertler A, Bolsi A, Hug EB. 2012. Spot scanning-based proton therapy for intracranial meningioma: Long-term results from the Paul Scherrer Institute. Int J Radiat Oncol Biol Phys 83:865–871.

Weller M, van den Bent M, Tonn JC, Stupp R, Preusser M, Cohen-Jonathan-Moyal E, Henriksson R, Rhun EL, Balana C, Chinot O, Bendszus M, Reijneveld JC, Dhermain F, French P, Marosi C, Watts C, Oberg I, Pilkington G, Baumert BG, Taphoorn MJB, Hegi M, Westphal M, Reifen-

berger G, Soffietti R, Wick W. 2017. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. Lancet Oncol 18:e315–e329.

Whittle IR, Smith C, Navoo P, Collie D. 2004. Meningiomas. Lancet 363:1535–1543.

Widder J, van der Schaaf A, Lambin P, Marijnen CA, Pignol JP, Rasch CR, Slotman BJ, Verheij M, Langendijk JA. 2016. The quest for evidence for proton therapy: Model-based approach and precision medicine. Int J Radiat Oncol Biol Phys 95:30–36.

Wilkinson B, Morgan H, Gondi V, Larson GL, Hartsell WF, Laramore GE, Halasz LM, Vargas C, Keole SR, Grosshans DR, Shih HA, Mehta MP. 2016. Low levels of acute toxicity associated with proton therapy for low-grade glioma: A proton collaborative group study. Int J Radiat Oncol Biol Phys 96:E135.

Willers H, Allen A, Grosshans D, McMahon SJ, von Neubeck C, Wiese C, Vikram B. 2018. Toward a variable RBE for proton beam therapy. Radiother Oncol 128:68–75.

Wilson CB. 1994. Meningiomas: genetics, malignancy, and the role of radiation in induction and treatment. J Neurosurg 81:666–675.

Wopken K, Bijl HP, van der Schaaf A, van der Laan HP, Chouvalova O, Steenbakkers RJ, Doornaert P, Slotman BJ, Oosting SF, Christianen ME, van der Laan BF, Roodenburg JL, Leemans CR, Verdonck-de Leeuw IM, Langendijk JA. 2014. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. Radiother Oncol 113:95–101.

Wouters BG, Begg AC. 2009. Irradiation-induced damage and the DNA damage response. In: Joiner MC, van der Kogel AJ (Eds) Basic Clinical Radiobiology. 4th ed. Hodder Education, London, pp. 11–26.

Wright JL, Yom SS, Awan MJ, Dawes S, Fischer-Valuck B, Kudner R, Mailhot Vega R, Rodrigues G. 2019. Standardizing normal tissue contouring for radiation therapy treatment planning: An ASTRO consensus paper. Pract Radiat Oncol 9:65–72.

Wu Q, Djajaputra D, Wu Y, Zhou J, Liu HH, Mohan R. 2003. Intensity-modulated radiotherapy optimization with gEUD-guided dose-volume objectives. Phys Med Biol 48:279–291.

Yang CC, Lin SY, Tseng CK. 2019. Maintenance of multidomain neurocognitive functions in pediatric patients after proton beam therapy: A prospective case-series study. Appl Neuropsychol Child 8:389–395.

Yoon M, Park SY, Shin D, Lee SB, Pyo HR, Kim DY, Cho KH. 2007. A new homogeneity index based on statistical analysis of the dose-volume histogram. J Appl Clin Med Phys 8:9–17.

Zhao R, Kong W, Shang J, Zhe H, Wang YY. 2017. Hippocampal-sparing whole-brain radiotherapy for lung cancer. Clin Lung Cancer 18:127–131.

Zips D, Baumann M. 2013. Place of proton radiotherapy in future radiotherapy practice. Semin Radiat Oncol 23:149–153.

Zureick AH, Evans CL, Niemierko A, Grieco JA, Nichols AJ, Fullerton BC, Hess CB, Goebel CP, Gallotto SL, Weyman EA, Gaudet DE, Nartowicz JA, Ebb DH, Jones RM, MacDonald SM, Tarbell NJ, Yock TI, Pulsifer MB. 2018. Left hippocampal dosimetry correlates with visual and verbal memory outcomes in survivors of pediatric brain tumors. Cancer 124:2238–2245.

Appendix

A Theoretical background

Table A.1: Scales and number of questions of the quality of life core questionnaire EORTC QLQ-C30.

Scale	Name	Number of questions
Global health status		
Global health status	QL	2
Functional scales		
Physical function	PF	6
Role function	RF	2
Emotional function	EF	4
Cognitive function	CF	3
Social function	SF	2
Symptom scales		
Fatigue	FA	3
Nausea and vomiting	NV	2
Pain	PA	2
Dyspnoea	DY	1
Insomnia	SL	1
Appetite loss	AP	1
Constipation	CO	1
Diarrhoea	DI	1
Financial difficulties	FI	1

Table A.2: Scales and number of questions of the brain-tumour specific quality of life questionnaire EORTC QLQ-BN20.

Scale	Name	Number of question
Future uncertainty	BNFU	4
Visual disorder	BNVD	3
Motor dysfunction	BNMD	3
Communication deficit	BNCD	3
Headaches	BNHA	1
Seizures	BNSE	1
Drowsiness	BNDR	1
Itchy skin	BNIS	1
Hair loss	BNHL	1
Weakness of legs	BNWL	1
Bladder control	BNBC	1

The Montreal Cognitive Assessment test

The MoCA test includes different tasks assessing several neurocognitive domains. The visuospatial and executive domain is tested by the *Trial Making test*, *Cube Copy test* and *Clock Drawing test*. To test the domain naming, the patient has to recognise three animals that are rarely seen in Western countries. In the *Digit Span test* the patient has to repeat five numbers in forward and three in backward order. Here, the ability of attention is assessed. Focused attention, i.e. concentration is tested by *Letter tapping*: the interviewer reads a list of letters and the patient has to tap his hand every time a certain letter is named. For calculation, the patient has to subtract 7 from 100. Two tests assess the language domain: *Sentence Repetition* and *Letter fluency*, for that the patient has to name as many words as possible beginning with a certain letter in one minute. For abstraction, similarity between objects has to be found. Delayed recall assesses the short-term memory. For that, the patient has to remember 5 words that were read by the interviewer during the MoCA test. To assess the ability of orientation in time and space the patient has to name the current date, month, year, day as well as place and city (Julayanont et al., 2013).

Table A.3: Analysis types of prediction models according to the TRIPOD classification (Moons et al., 2015).

Type	Description
1a	Development of a predictive model and direct evaluation of model performance on the same data set
1b	Development of a predictive model using the entire data set including resampling strategies
2a	Random split of the cohort in two groups, a training and a validation cohort
2b	Nonrandom split of the cohort in two groups, a training and a validation cohort
3	Development of a predictive model on a training cohort and evaluation on a separate cohort
4	Evaluation of the performance of an existing predictive model on a separate cohort

B Modelling of side effects following cranial proton beam therapy

Table B.4: Clinical goals for treatment planning in the cohort treated at UPTD. Given are dose-volume parameters of different organs at risk and the values, which should not be exceeded.

Organ at risk	Parameter	Value	
Brain stem	D_{max}	54 Gy(RBE)	(Mayo et al., 2010b)
Spinal cord	D_{max}	45 – 60 Gy(RBE)	
Inner ear	D_{mean}	45 – 60 Gy(RBE)	(Jereczek-Fossa et al., 2003)
Chiasma	D_{max}	54 Gy(RBE)	(Mayo et al., 2010a)
Optic nerves	D_{max}	54 Gy(RBE)	(Mayo et al., 2010a)
Lenses	D_{max}	5 – 6 Gy(RBE)	
Lacrimal glands	D_{max}	40 – 45 Gy(RBE)	(Gordon et al., 1995)
Salivary glands	D_{median}	26 Gy(RBE)	
Brain	D_{mean}	40 Gy(RBE)	(Lawrence et al., 2010)

Table B.5: Clinical goals for treatment planning in the cohort treated at WPE. Given are dose-volume parameters of different organs at risk and the values, which should not be exceeded.

Organ at risk	Chordoma, chondrosarcoma		Meningeoma I		Meningeoma II–III		Glioma II		Glioma III–IV	
Brain stem	D_{max}	66 Gy(RBE)	D_{max}	55 Gy(RBE)	D_{max}	60 Gy(RBE)	D_{max}	55 Gy(RBE)	D_{max}	60 Gy(RBE)
	$D1\%$	64 Gy(RBE)								
	$D2\%$	62 Gy(RBE)								
Brain stem, centre	$D2\%$	54 Gy(RBE)			D_{max}	54 Gy(RBE)				
Spinal cord	D_{max}	54 Gy(RBE)								
	D_{mean}	50 Gy(RBE)								
Spinal cord, centre	$D2\%$	50 Gy(RBE)								
Inner ear, ipsilateral	D_{max}	$< 105\,\%$								
Inner ear, at least one side	D_{max}	60 Gy(RBE)								
Cochlea	D_{mean}	45 Gy(RBE)	D_{mean}	45 Gy(RBE)	D_{mean}	45 Gy(RBE)	D_{mean}	45 Gy(RBE)	D_{mean}	45 Gy(RBE)
Chiasma	D2%	60 Gy(RBE)	D_{max}	56 Gy(RBE)	D_{max}	60 Gy(RBE)	D_{max}	55 Gy(RBE)	D_{max}	60 Gy(RBE)
Optic nerves	$D2\%$	60 Gy(RBE)	D_{max}	56 Gy(RBE)	D_{max}	56–60 Gy(RBE)	D_{max}	55 Gy(RBE)	D_{max}	60 Gy(RBE)
Temporal lobes	$D2\%$	70 Gy(RBE)								
	$V60Gy(RBE)$	$< 10\,\%$								
Lenses	D_{mean}	10 Gy(RBE)								
Lacrimal glands	D_{mean}	36 Gy(RBE)	D_{mean}	36 Gy(RBE)	D_{mean}	36 Gy(RBE)	D_{mean}	36 Gy(RBE)	D_{mean}	36 Gy(RBE)
Salivary gland, unilateral	$D2\%$	36 Gy(RBE)								
Salivary gland, contralateral	D_{mean}	26 Gy(RBE)								
Brain, contralateral			D_{max}	36 Gy(RBE)	D_{max}	36 Gy(RBE)	D_{max}	36 Gy(RBE)	D_{max}	36 Gy(RBE)

Table B.6: Results of univariable logistic regression for clinical cofactors and early side effects.

Clinical cofactor		Erythema grade ≥1	Erythema grade ≥2	Alopecia grade ≥1	Alopecia grade ≥2	Fatigue grade ≥1
Surgery	p-value	< 0.001	0.085	0.36	0.030	0.40
No/yes	OR	17.52	3.90	1.93	3.66	0.56
	(95 % CI)	(4.74 – 64.82)	(0.83 – 18.38)	(0.47 – 7.93)	(1.13 – 11.81)	(0.15 – 2.16)
Gender	p-value	0.094	0.90	0.96	0.15	0.005
Male/female	OR	0.39	0.95	1.03	1.81	3.90
	(95 % CI)	(0.13 – 1.18)	(0.43 – 2.08)	(0.34 – 3.12)	(0.81 – 4.07)	(1.52 – 9.99)
Tumour location	p-value	0.68	0.83	0.83	0.77	0.90
Brain/skull base	OR	1.29	1.09	0.88	1.13	0.95
	(95 % CI)	(0.38 – 4.38)	(0.48 – 2.51)	(0.28 – 2.80)	(0.49 – 2.63)	(0.40 – 2.25)
Chemotherapy	p-value	0.49	0.44	0.47	0.29	0.75
No/yes	OR	1.60	1.39	1.65	1.63	1.16
	(95 % CI)	(0.42 – 6.10)	(0.60 – 3.25)	(0.43 – 6.29)	(0.67 – 3.99)	(0.47 – 2.86)
Age at PBT	p-value	0.85	0.43	0.29	0.14	0.12
in years	OR	1.00	0.99	0.98	0.98	1.02
	(95 % CI)	(0.96 – 1.03)	(0.96 – 1.02)	(0.94 – 1.02)	(0.95 – 1.01)	(0.99 – 1.05)
Tumour volume (CTV)	p-value	0.004	0.024	< 0.001	< 0.001	0.82
in cm^3	OR	1.01	1.00	1.03	1.01	1.00
	(95 % CI)	(1.00 – 1.02)	(1.00 – 1.01)	(1.01 – 1.04)	(1.01 – 1.02)	(1.00 – 1.00)
Prescribed total dose	p-value	0.037	0.001	0.028	0.033	0.84
in Gy(RBE)	OR	1.07	1.14	1.07	1.06	1.01
	(95 % CI)	(1.00- 1.14)	(1.05 – 1.23)	(1.01 – 1.14)	(1.00 – 1.12)	(0.95 – 1.06)

Abbreviations: OR, Odds Ratio; CI, confidence interval; CTV, clinical target volume.

Table B.7: Comparison of selected dose-volume parameters of the skin between the exploration and validation cohort 1 for different total dose prescriptions (54 Gy(RBE) and 60 Gy(RBE)). P-values represent results from the Mann-Whitney U test.

	Prescribed total dose 54 Gy(RBE)					Prescribed total dose 60 Gy(RBE)				
		Exploration N = 26		Validation 1 N = 43			Exploration N = 56		Validation 1 N = 19	
Parameter in cm^3 or Gy(RBE)	p-value	Mean	SD	Mean	SD	p-value	Mean	SD	Mean	SD
$V5Gy(RBE)$	0.001	34	19	57	36	0.044	57	18	74	30
$V10Gy(RBE)$	0.11	30	17	43	33	0.10	53	17	64	27
$V20Gy(RBE)$	0.061	21	16	19	29	0.81	46	16	47	27
$V30Gy(RBE)$	0.15	6	11	13	23	0.29	28	16	33	23
$V40Gy(RBE)$	0.34	4	8.1	8	17	0.29	17	12	23	19
$V50Gy(RBE)$	0.84	2.1	4.5	2.7	6.6	0.78	8.5	8.6	8.9	9.9
$V60Gy(RBE)$						0.01	0.7	2.2	0.00	0.01
$D2\%$	0.20	25	15	23	15	0.14	44	11	39	14
$D5\%$	0.86	16	15	14	12	0.079	34	12	26	15
$D15\%$	0.012	0.8	1.9	2.0	5.6	0.24	5.3	8.7	1.7	3.2
$D25\%$	< 0.001	0.02	0.04	0.29	0.87	0.31	0.5	2.9	0.11	0.22
$D35\%$	< 0.001	0	0.02	0.06	0.19	< 0.001	0.02	0.09	0.04	0.04
$D45\%$	< 0.001	0	0	0.02	0.03	< 0.001	0.00	0.01	0.02	0.01

Abbreviations: N, number of patients; SD, standard deviation.

Table B.8: Spearman correlation coefficients ρ for clinical cofactors and dose-volume parameters selected as predictors for different acute side-effects. Tumour location determines whether the tumour is located in the brain or the skull base.

		Surgery [no/yes]	Gender [male/female]	Location [brain/skull base]	CTx [no/yes]	Age in years	CTV in cm^3	Prescribed dose in Gy(RBE)
Skin *V*25*Gy*(*RBE*) in cm^3	ρ	0.38	-0.05	-0.08	0.09	0.03	0.83	0.70
	p-value	< 0.001	0.59	0.41	0.32	0.73	< 0.001	< 0.001
Skin *V*35*Gy*(*RBE*) in cm^3	ρ	0.41	-0.05	-0.03	0.05	0.01	0.75	0.69
	p-value	< 0.001	0.57	0.76	0.58	0.94	< 0.001	< 0.001
Skin *D*2% in Gy(RBE)	ρ	0.38	0.03	-0.10	0.07	0.00	0.74	0.65
	p-value	< 0.001	0.77	0.28	0.45	0.99	< 0.001	< 0.001
Skin *D*5% in Gy(RBE)	ρ	0.34	0.06	-0.04	0.21	0.04	0.81	0.50
	p-value	< 0.001	0.52	0.69	0.03	0.67	< 0.001	< 0.001
Brain-CTV *D*2% in Gy(RBE)	ρ	0.15	0.15	0.02	0.02	-0.06	0.59	0.31
	p-value	0.11	0.11	0.84	0.83	0.53	< 0.001	0.001
Surgery [no/yes]	ρ	1.00	-0.19	-0.09	0.05	0.02	0.31	0.24
	p-value		0.046	0.32	0.59	0.80	0.001	0.010
Gender [male/female]	ρ	-0.19	1.00	0.01	-0.10	-0.04	-0.14	-0.03
	p-value	0.046		0.92	0.32	0.70	0.14	0.79
Location [brain/skull base]	ρ	-0.09	0.01	1.00	-0.15	-0.17	-0.06	-0.06
	p-value	0.32	0.92		0.12	0.08	0.51	0.50
CTx [no/yes]	ρ	0.05	-0.10	-0.15	1.00	0.02	0.37	0.09
	p-value	0.59	0.32	0.12		0.85	< 0.001	0.32
Age at PBT [years]	ρ	0.02	-0.04	-0.17	0.02	1.00	-0.04	0.00
	p-value	0.80	0.70	0.08	0.85		0.71	0.99
CTV in cm^3	ρ	0.31	-0.14	-0.06	0.37	-0.04	1.00	0.55
	p-value	0.001	0.14	0.51	< 0.001	0.71		< 0.001
Prescribed dose in Gy(RBE)	ρ	0.24	-0.03	-0.06	0.09	0.00	0.55	1.00
	p-value	0.010	0.79	0.50	0.32	0.99	< 0.001	

Abbreviations: CTx, chemotherapy; CTV, clinical target volume.

Table B.9: Results of the principal component analysis. The first three unrotated and rotated components with eigenvalues greater than one are given and their corresponding explained total variance.

		Percentage of total variance		Dose-volume parameters represented by
Component	**Eigenvalue**	**Without rotation**	**Varimax rotation**	**rotated components**
Component 1	11.3	70.8%	49.3%	$V10Gy(RBE) - V20Gy(RBE)$, $D2\%$, $D5\%$
Component 2	1.9	12.0%	30.7%	$V50Gy(RBE) - V60Gy(RBE)$
Component 3	1.1	7.0%	9.9%	$D15\%$, $D25\%$

Table B.10: Univariable logistic modelling results: Rotated (Varimax) principal components of the skin for acute erythema and alopecia. Mean AUC values for 3-fold cross-validation (333 repetitions) and external validation are given. P-values were calculated on the exploration cohort, confidence intervals (CI) were obtained using 1000 bootstrap samples. Fitting parameters β_i as defined in equation (2.13).

	Cross validation		External validation cohort 1		External validation cohort 2						
Model	AUC	(95% CI)	AUC	(95 % CI)	AUC	(95 % CI)	β_0	(95 % CI)	β_1	(95 % CI)	p-value
Erythema grade $\geq$ 1											
Component 1	0.75	(0.58 – 0.89)	0.81	(0.69 – 0.91)	0.76	(0.65 – 0.86)	2.12	(1.46 - 2.78)	0.83	(0.23 – 1.42)	0.006
Erythema grade $\geq$ 2											
Component 1	0.69	(0.55 – 0.82)	*		0.00	(0.00 – 0.00)	-0.63	(-1.04 – 0.22)	0.70	(0.25 - 1.15)	0.002
Alopecia grade $\geq$ 1											
Component 1	0.85	(0.64 – 0.99)	0.83	(0.71 – 0.92)	0.83	(0.73 – 0.92)	2.80	(1.77 – 3.83)	1.71	(0.89 – 2.53)	$<$ 0.001
Alopecia grade $\geq$ 2											
Component 1	0.83	(0.71 – 0.94)	0.75	(0.60 – 0.87)	0.88	(0.81 – 0.94)	0.74	(0.25 – 1.22)	1.51	(0.94 – 2.09)	$<$ 0.001

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval. *not applicable, as the incidence rate was zero

Table B.11: Patient characteristics for the generalised estimating equation analysis of late side effects for the pooled cohort.

Total number of patients	216	
Characteristics	**Median**	**(Range)**
Age[1] in years	47.8	(18.1 – 89.3)
Tumour volume (CTV)[1] in cm^3	51.4	(1.2 – 498.0)
Prescribed dose[1] in Gy(RBE)	54.0	(30.0 – 74.0)
	N	**(%)**
Gender[2] (male/female)	103/113	(48/52)
Surgery[2] (no/yes/missing)	44/171/1	(20/79/0.5)
Chemotherapy[2] (no/yes/missing)	157/58/1	(73/27/0.5)
Re-irradiation[2] (no/yes/missing)	201/13/2	(93/6/1)
Tumour location[2] (brain/skull base/other)	99/113/4	(46/52/2)
Tumour location[2] (hemisphere)		
Left	74	(34)
Right	81	(38)
Central	54	(25)
Bilateral	7	(3)
Tumour location[2] (lobe)		
Temporal lobe	66	(31)
Frontal lobe	51	(24)
Parietal lobe	11	(5)
Occipital lobe	5	(2)
Multiple lobes	30	(14)
Other	53	(25)
Tumour histology[*,2]		
High-grade glioma	51	(24)
Low-grade glioma	30	(14)
Meningioma	67	(31)
Other	68	(31)

Abbreviations: N, number of patients; CTV, clinical target volume. *Tumour classification according to WHO Classification of Tumours of the Central Nervous System. Fourth Edition. [1]Mann-Whitney U test, [2] χ^2 test.

Table B.12: Generalised estimating equation analyses for late side effects and cofactors in the pooled analysis. The parameter time is defined as a natural number representing the follow-up visit starting with 1 at 6 months following PBT.

Excluding interaction term between dosimetric parameter and time				Including interaction term between dosimetric parameter and time			
Model parameter	β_i	**(95 % CI)**	**p-value**	**Model parameter**	β_i	**(95 % CI)**	**p-value**
Alopecia grade ≥ 1							
Prescribed dose in Gy(RBE)$^{-1}$	0.29	(0.16 – 0.42)	< 0.001	Prescribed dose in Gy(RBE)$^{-1}$	0.27	(0.09 – 0.44)	0.003
Time	-0.031	(-0.06 – 0.00)	0.046	Time	-0.13	(-0.74 – 0.48)	0.67
Constant	-16.97	(-24.50 – -9.44)		Interaction	0.002	(-0.01 – 0.01)	0.74
				Constant	-15.54	(-25.61 – -5.47)	
Tumour volume (CTV) in cm^{-3}	0.009	(0.01 – 0.01)	< 0.001	Tumour volume (CTV) in cm^{-3}	0.008	(0.00 – 0.01)	0.017
Time	-0.030	(-0.06 – 0.00)	0.058	Time	-0.044	(-0.10 – 0.01)	0.094
Constant	-1.56	(-2.17 – -0.95)		Interaction	0.000	(0.00 – 0.00)	0.53
				Constant	-1.36	(-2.22 – -0.50)	
Alopecia grade ≥ 2							
Prescribed dose in Gy(RBE)$^{-1}$	0.25	(0.09 – 0.41)	0.002	Prescribed dose in Gy(RBE)$^{-1}$	0.39	(0.05 – 0.73)	0.025
Time	-0.013	(-0.10 – 0.08)	0.78	Time	0.52	(-0.38 – 1.41)	0.26
Constant	-17.93	(-27.01 – -8.85)		Interaction	-0.009	(-0.02 – 0.01)	0.24
				Constant	-26.08	(-45.95 – -6.21)	
Tumour volume (CTV) in cm^{-3}	0.007	(0.00 – 0.01)	< 0.001	Tumour volume (CTV) in cm^{-3}	0.011	(0.00 – 0.02)	< 0.001
Time	-0.012	(-0.10 – 0.08)	0.78	Time	0.038	(-0.10 – 0.17)	0.58
Constant	-4.42	(-5.88 – -2.96)		Interaction	0.000	(0.00 – 0.00)	0.22
				Constant	-5.17	(-7.26 – -3.08)	
Dry eye syndrome grade ≥ 2							
Tumour located in brain [ref: skull base]	-2.78	(-4.93 – -0.63)	0.011	Tumour located in brain [ref: skull base]	-7.11	(-9.37 – -4.84)	< 0.001
Time	-0.040	(-0.12 – 0.04)	0.32	Time	-0.056	(-0.14 – 0.03)	0.19
Constant	-2.49	(-3.81 – -1.17)		Interaction	0.25	(0.17 – 0.34)	< 0.001
				Constant	-2.29	(-3.51 – -1.06)	
Headache grade ≥ 1							
Age at PBT in 1/years	-0.021	(-0.04 – -0.01)	0.010	Age at PBT in 1/years	0.008	(-0.02 – 0.03)	0.55
Time	-0.019	(-0.04 – 0.01)	0.13	Time	0.077	(0.00 – 0.15)	0.040
Constant	0.083	(-0.75 – 0.92)		Interaction	-0.002	(0.00 – 0.00)	0.012
				Constant	-1.21	(-2.41 – -0.02)	
Hearing impairment grade ≥ 1							
Age at PBT in 1/years	0.057	(0.02 – 0.09)	0.001	Age at PBT in 1/years	0.057	(0.00 – 0.11)	0.034
Time	0.006	(-0.04 – 0.05)	0.79	Time	0.006	(-0.16 – 0.17)	0.94
Constant	-5.40	(-7.55 – -3.26)		Interaction	0.000	(0.00 – 0.00)	0.99
				Constant	-5.41	(-8.56 – -2.26)	
Memory impairment grade ≥ 1							
Prescribed dose in Gy(RBE)$^{-1}$	0.064	(0.01 – 0.11)	0.011	Prescribed dose in Gy(RBE)$^{-1}$	0.091	(0.02 – 0.16)	0.012
Time	0.020	(-0.01 – 0.05)	0.16	Time	0.130	(-0.08 – 0.34)	0.23
Constant	-5.20	(-7.96 – -2.44)		Interaction	-0.002	(-0.01 – 0.00)	0.31
				Constant	-6.78	(-10.79 – -2.77)	
Tumour volume (CTV) in cm^{-3}	0.004	0.00 0.01	0.004	Tumour volume (CTV) in cm^{-3}	0.005	(0.00 0.01	0.042
Time	0.021	(-0.01 – 0.05)	0.15	Time	0.029	(-0.01 – 0.07)	0.16
Constant	-2.14	(-2.75 – -1.53)		Interaction	0.000	(0.00 – 0.00)	0.63
				Constant	-2.27	(-3.05 – -1.48)	

Abbreviations: CI, confidence interval; CTV, clinical target volume.

Table B.13: Definitions of alopecia and dermatitis according to CTCAE v4.0 (National Cancer Institute, 2009).

Alopecia	
A disorder characterized by a decrease in density of hair compared to normal for a given individual at a given age and body location.	
Grade 0	Absence of alopecia or within normal limits.
Grade 1	Hair loss of $<$ 50 % of normal for that individual that is not obvious from a distance but only on close inspection; a different hairstyle may be required to cover the hair loss but it does not require a wig or hair piece to camouflage.
Grade 2	Hair loss of $\geq$ 50 % normal for that individual that is readily apparent to others; a wig or hair piece is necessary if the patient desires to completely camouflage the hair loss; associated with psychosocial impact.
Radiation dermatitis	
A finding of cutaneous inflammatory reaction occurring as a result of exposure to biologically effective levels of ionizing radiation.	
Grade 0	Absence of dermatitis or within normal limits.
Grade 1	Faint erythema or dry desquamation.
Grade 2	Moderate to brisk erythema; patchy moist desquamation, mostly confined to skin folds and creases; moderate oedema.
Grade 3	Moist desquamation in areas other than skin folds and creases; bleeding induced by minor trauma or abrasion.
Grade 4	Life-threatening consequences; skin necrosis or ulceration of full-thickness dermis; spontaneous bleeding from the involved site; skin graft indicated.
Grade 5	Death

Figure B.1: Example case of the interobserver variability study.

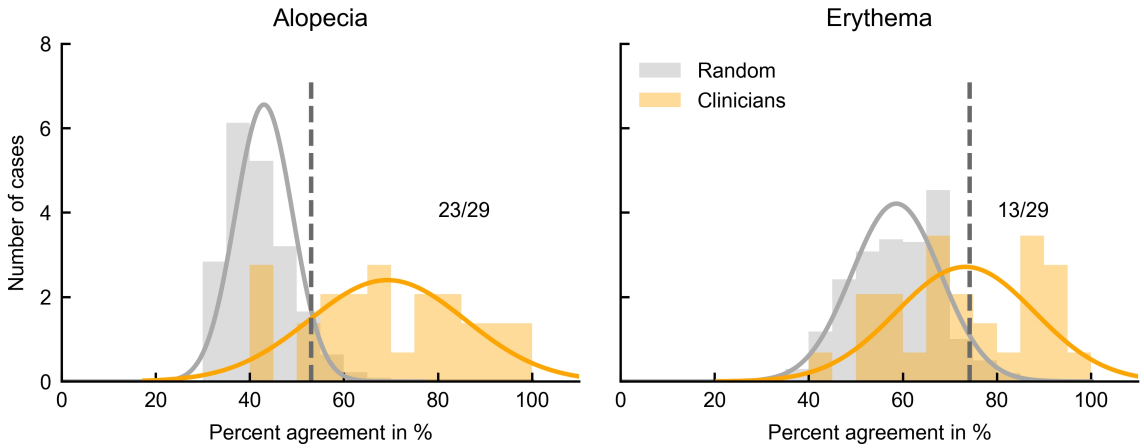


Figure B.2: Percent agreement for scoring of alopecia and erythema compared to random agreement based on the distribution of the modal value. The 95 % percentile is given as dashed line. The number of cases above this threshold out of the number of total cases (29) is indicated.

C Neurocognitive function following proton beam therapy

Table C.1: Response rates and scores of the MoCA test and quality of life questionnaires EORTC QLQ-C30 and EORTC QLQ-BN20 over time (months after treatment are given). Data are given in mean (standard deviation). MoCA score ranges from 0 to 30, QoL items from 0 to 100.

	Baseline	3 months	6 months	9 months	12 months	15 months	18 months	21 months	24 months
MoCA test									
Number of patients	62	48	38	35	27	27	21	24	26
MoCA total score	24.8 (4.2)	24.7 (4.2)	25.1 (4.1)	25.9 (3.6)	25.7 (3.3)	26.7 (3.5)	24.8 (4.9)	26.4 (2.9)	25.2 (4.2)
EORTC-QLQ-C30									
Number of patients	61	48	37	37	30	30	25	26	26
Cognitive function	68.9 (26.6)	64.2 (30.0)	63.5 (29.1)	65.3 (29.2)	66.1 (30.8)	71.1 (23.5)	62.7 (29.8)	64.1 (27.4)	66.0 (25.2)
Physical function	75.4 (23.8)	73.5 (25.1)	75.4 (24.3)	76.0 (25.4)	76.0 (23.7)	82.0 (15.6)	77.3 (24.6)	79.5 (17.9)	75.1 (22.0)
Social function	61.7 (32.5)	57.6 (35.7)	58.6 (38.2)	66.2 (31.5)	57.5 (33.5)	70.0 (26.8)	57.3 (33.5)	59.6 (34.3)	60.3 (32.0)
Emotional function	61.2 (25.4)	62.7 (25.4)	65.5 (25.7)	65.3 (25.6)	61.3 (2.38)	69.2 (19.6)	58.6 (25.5)	64.1 (24.2)	66.0 (23.4)
Role function	64.7 (34.3)	60.4 (36.2)	64.5 (32.5)	69.4 (33.9)	60.6 (36.0)	64.4 (29.9)	58.0 (35.7)	60.3 (31.3)	60.3 (34.0)
Global health status	58.9 (22.2)	58.0 (26.3)	58.3 (23.9)	64.6 (23.7)	56.9 (26.0)	65.6 (21.1)	54.0 (26.1)	56.4 (26.6)	56.1 (23.9)
Fatigue	43.0 (29.6)	46.1 (30.4)	48.2 (30.5)	39.3 (26.7)	48.9 (33.3)	38.5 (28.8)	48.9 (33.3)	38.5 (28.8)	48.5 (28.1)
Pain	27.0 (29.2)	32.3 (35.1)	28.9 (29.7)	27.5 (31.7)	31.7 (35.4)	26.1 (31.2)	32.7 (31.0)	23.1 (29.5)	37.2 (32.8)
Insomnia	27.9 (33.4)	26.4 (32.2)	31.6 (30.9)	34.2 (32.9)	34.4 (39.6)	27.8 (34.0)	37.3 (35.1)	30.8 (31.2)	28.2 (29.4)
EORTC-QLQ-BN20									
Number of patients	61	48	39	38	30	29	26	26	25
Communication deficit	20.4 (26.3)	20.4 (26.1)	26.3 (30.2)	23.4 (27.2)	21.5 (25.3)	15.7 (20.9)	24.0 (27.2)	18.8 (22.6)	20.4 (21.7)
Visual disorder	21.1 (23.3)	16.4 (22.1)	15.5 (21.1)	19.5 (25.8)	17.4 (24.0)	19.9 (30.0)	20.9 (24.3)	20.9 (28.7)	27.6 (32.6)
Future uncertainty	44.7 (26.3)	38.5 (27.4)	35.9 (28.1)	32.4 (26.5)	36.9 (27.4)	28.4 (24.9)	44.7 (30.7)	34.9 (25.7)	33.7 (24.6)
Headache	36.6 (32.6)	33.3 (35.1)	28.9 (30.2)	32.4 (34.7)	38.9 (35.1)	31.0 (34.4)	34.7 (32.6)	26.9 (31.3)	34.7 (34.7)
Seizures	2.7 (14.0)	5.6 (15.9)	7.6 (23.0)	1.9 (11.1)	1.1 (6.2)	6.0 (22.3)	6.7 (13.6)	2.6 (13.1)	5.3 (18.5)
Hair loss	8.8 (19.4)	26.2 (34.0)	23.7 (33.7)	24.3 (34.8)	25.6 (36.8)	13.8 (30.2)	16.0 (27.4)	16.7 (33.0)	18.1 (35.4)

$^{+}$high score: high functionality, $^{-}$high score: high symptomatology.

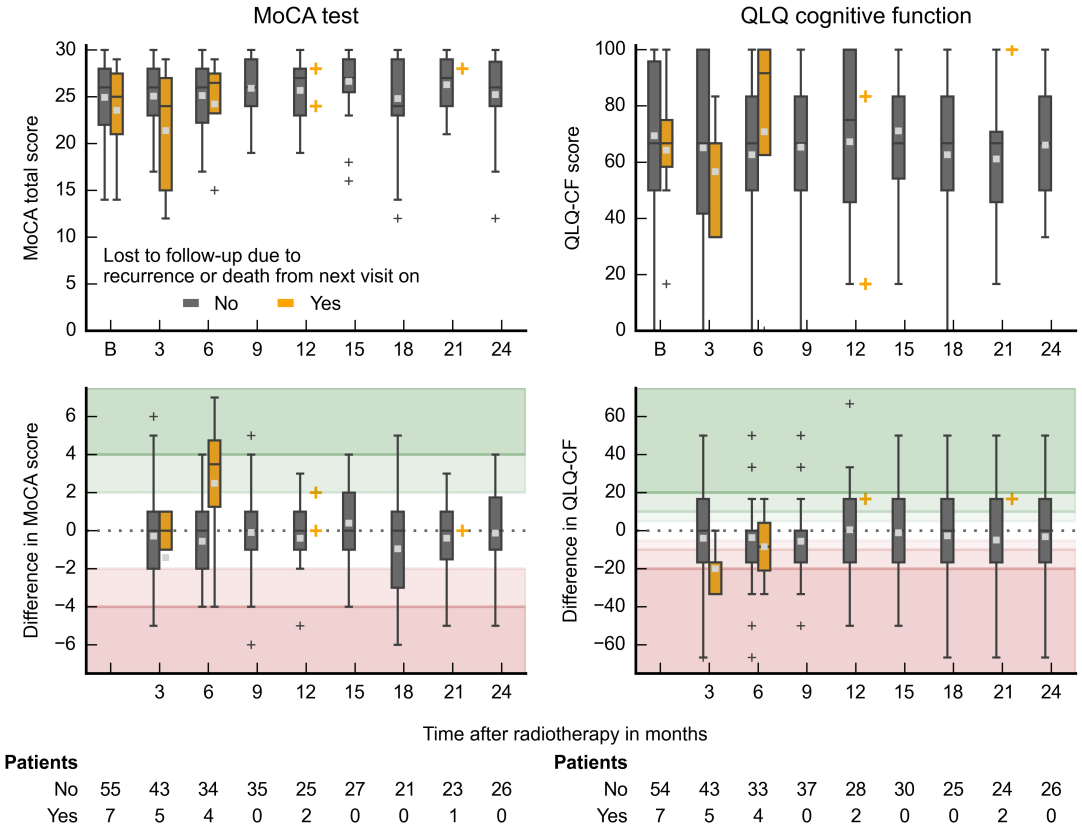


Figure C.1: MoCA total score and self-reported cognitive function (QLQ cognitive function) over time. Patients are classified whether they were lost to follow-up from the next visit on due to death or local recurrence or not. Absolute scores in the top row, differences to baseline (B) below. Shaded areas mark the extent of changes in QLQ and MoCA (red: worsening, green: improvement). Extent of QLQ changes from light to dark: a little (5.0 – 9.9), moderate (10.0 – 19.9) and very much ($\geq$ 20.0) (Osoba et al., 1998). Extent of MoCA changes from light to dark: moderate (2 – 4 points), large ($\geq$ 4 points). No significant differences between both groups could be observed at any time. Adapted from Dutz et al. (2020).

Table C.2: Results of the mixed model analyses on differences from baseline for selected quality of life items and clinical parameters including all times. Model coefficients and standard deviation (standard deviation) of the parameters, time after proton beam therapy and their interaction term are given.

Model parameter	Coefficient (SD)	p-value
QLQ-C30 Physical function		
Corticosteroids [reference: yes]	13.4 (6.00)	0.028
Time after radiotherapy	0.72 (0.93)	0.44
Interaction of time and corticosteroids	-1.17 (1.14)	0.31
QLQ-C30 Emotional function		
Age at PBT in years	-0.63 (0.27)	0.022
Time after radiotherapy	-2.81 (2.42)	0.25
Interaction of time and age	0.06 (0.05)	0.23
QLQ-C30 Social function		
Age at PBT in years	-0.83 (0.35)	0.019
Time after radiotherapy	-1.42 (3.10)	0.65
Interaction of time and age	0.03 (0.06)	0.64
QLQ-C30 Role function		
Gender [reference: female]	38.4 (11.1)	0.001
Time after radiotherapy	1.45 (1.57)	0.36
Interaction of time and gender	-2.66 (2.04)	0.19
Corticosteroids [reference: yes]	24.3 (11.8)	0.042
Time after radiotherapy	1.62 (1.86)	0.38
Interaction of time and corticosteroids	-2.83 (2.28)	0.22
QLQ-C30 Pain		
Age at PBT in years	0.63 (0.26)	0.019
Time after radiotherapy	1.86 (2.59)	0.47
Interaction of time and age	-0.03 (0.05)	0.53
QLQ-BN20 Visual disorder		
Surgery [reference: yes]	-13.1 (8.55)	0.13
Time after radiotherapy	-0.02 (0.65)	0.97
Interaction of time and surgery	4.70 (1.63)	0.005
QLQ-BN20 Seizures		
Corticosteroids [reference: yes]	-11.0 (4.47)	0.015
Time after radiotherapy	-1.17 (0.77)	0.14
Interaction of time and corticosteroids	1.47 (0.95)	0.13

Abbreviations: QLQ-BN20, brain tumour-specific quality of life questionnaire; QLQ-C30, quality of life core questionnaire; SD, standard deviation.

Table C.3: Results of the mixed model analyses on differences from baseline of the MoCA total score and self-reported cognitive function and dose-volume parameters of the hippocampus including all time points. Model coefficients and standard deviation (standard deviation) of dosimetric parameters, time after proton beam therapy and their interaction term are given.

	MoCA test		QLQ cognitive function	
Model parameter	**Coefficient (SD)**	**p-value**	**Coefficient (SD)**	**p-value**
Hippocampi $V10Gy(RBE)$	1.45 (1.73)	0.40	0.66 (16.6)	0.97
Time	0.04 (0.13)	0.75	0.55 (1.26)	0.67
Interaction of time and hippocampi $V10Gy(RBE)$	-0.15 (0.33)	0.66	-1.85 (3.17)	0.56
Constant	-0.85 (0.66)	0.20	-5.4 (6.34)	0.40
Hippocampi $V20Gy(RBE)$	1.52 (1.73)	0.38	-0.78 (16.5)	0.96
Time	0.04 (0.12)	0.76	0.56 (1.18)	0.64
Interaction of time and hippocampi $V20Gy(RBE)$	-0.14 (0.33)	0.66	-2.08 (3.14)	0.51
Constant	-0.83 (0.62)	0.19	-5.02 (5.94)	0.40
Hippocampi $V30Gy(RBE)$	1.59 (1.76)	0.37	-1.14 (16.8)	0.95
Time	0.04 (0.12)	0.75	0.51 (1.14)	0.65
Interaction of time and hippocampi $V30Gy(RBE)$	-0.17 (0.34)	0.63	-2.10 (3.19)	0.51
Constant	-0.81 (0.60)	0.18	-4.96 (5.70)	0.39
Hippocampi $V40Gy(RBE)$	1.74 (1.83)	0.35	2.23 (17.6)	0.90
Time	0.03 (0.12)	0.79	0.52 (1.11)	0.64
Interaction of time and hippocampi $V40Gy(RBE)$	-0.16 (0.36)	0.67	-2.45 (3.41)	0.47
Constant	-0.79 (0.58)	0.17	-5.64 (5.48)	0.31
Hippocampi V50Gy(RBE)	2.92 (2.17)	0.18	3.86 (21.2)	0.86
Time	0.01 (0.11)	0.90	0.44 (1.09)	0.69
Interaction of time and hippocampi $V50Gy(RBE)$	-0.08 (0.45)	0.85	-2.86 (4.27)	0.50
Constant	-0.91 (0.56)	0.11	-5.76 (5.43)	0.29
Hippocampi $V60Gy(RBE)$	-4.59 (10.2)	0.65	-65.3 (103.2)	0.53
Time	-0.01 (0.10)	0.92	-0.04 (0.91)	0.96
Interaction of time and hippocampi $V60Gy(RBE)$	0.94 (2.03)	0.64	3.64 (18.9)	0.85
Constant	-0.38 (0.48)	0.43	-4.45 (4.55)	0.33
Hippocampi $D2\%$	0.02 (0.02)	0.34	0.01 (0.18)	0.96
Time	-0.08 (0.16)	0.62	1.00 (1.54)	0.52
Interaction of time and hippocampi $D2\%$	0.00 (0.00)	0.54	-0.03 (0.04)	0.44
Constant	-1.06 (0.77)	0.17	-5.56 (7.53)	0.46
Hippocampi D_{mean}	0.03 (0.03)	0.35	0.02 (0.33)	0.95
Time	0.04 (0.13)	0.77	0.56 (1.20)	0.64
Interaction of time and hippocampi D_{mean}	0.00 (0.01)	0.68	-0.04 (0.06)	0.51
Constant	-0.86 (0.63)	0.17	-5.44 (6.01)	0.37
Hippocampi $D50\%$	0.05 (0.03)	0.16	0.00 (0.32)	1.00
Time	0.05 (0.10)	0.62	0.13 (0.96)	0.89
Interaction of time and hippocampi $D50\%$	-0.01 (0.01)	0.23	-0.02 (0.06)	0.73
Constant	-0.71 (0.49)	0.15	-5.21 (4.76)	0.28

Abbreviation: SD, standard deviation.

Table C.4: Results of the mixed model analyses on score differences of the MoCA subscales compared to baseline including all times. Model coefficients (standard deviation) of the clinical or dosimetric parameters, time after proton beam therapy and their interaction term are given. Only parameters with statistically significant association are shown. If more than one dosimetric parameter of a structure was significantly associated to changes in the MoCA subscales score, the parameter with the lowest p-value is given.

Model parameter	Coefficient (SD)	p-value
MoCA test: memory		
Cerebellum $D2\%$	-0.02 (0.01)	0.041
Time	-0.06 (0.07)	0.41
Interaction of time and cerebellum $D2\%$	0.00 (0.00)	0.050
Constant	0.57 (0.33)	
Cerebellum anterior $V40Gy(RBE)$	-4.80 (1.67)	0.005
Time	-0.04 (0.06)	0.52
Interaction of time and cerebellum anterior $V40Gy(RBE)$	0.76 (0.33)	0.023
Constant	0.55 (0.28)	
Chemotherapy [reference: yes]	-1.48 (0.50)	0.004
Time	-0.05 (0.09)	0.56
Interaction of time and chemotherapy [reference: yes]	0.14 (0.11)	0.21
Constant	1.10 (0.41)	
MoCA test: orientation		
Right frontal lobe $V10Gy(RBE)$	-1.44 (0.60)	0.019
Time	0.02 (0.04)	0.51
Interaction of time and right frontal lobe $V10Gy(RBE)$	0.04 (0.11)	0.74
Constant	0.12 (0.19)	
Cerebellum $D50\%$	-0.06 (0.02)	0.018
Time	0.01 (0.02)	0.54
Interaction of time and cerebellum $D50\%$	0.01 (0.00)	0.27
Constant	-0.10 (0.13)	
Cerebellum posterior $D50\%$	-0.05 (0.02)	0.044
Time	0.02 (0.02)	0.52
Interaction of time and cerebellum posterior $D50\%$	0.00 (0.00)	0.33
Constant	-0.12 (0.13)	
MoCA test: attention		
ECOG performance status [reference: ECOG 2]		
ECOG 0	1.52 (0.84)	0.070
ECOG 1	2.09 (0.86)	0.016
Time	0.30 (0.23)	0.20
Interaction of time and ECOG 0	-0.26 (0.23)	0.26
Interaction of time and ECOG 1	-0.35 (0.23)	0.13
Constant	-1.83 (0.82)	
MoCA test: abstraction		
Cerebellum posterior $D50\%$	-0.02 (0.01)	0.035
Time	-0.02 (0.01)	0.22
Interaction of time and cerebellum posterior $D50\%$	0.00 (0.00)	0.10
Constant	0.09 (0.06)	
Age at therapy	0.01 (0.00)	0.010
Time	0.02 (0.04)	0.57
Interaction of time and age at therapy	0.00 (0.00)	0.40
Constant	-0.40 (0.18)	

Abbreviations: SD, standard deviation; ECOG, Eastern Co-operative of Oncology Group.

Table C.5: Mean baseline values for MoCA and quality of life values compared to reference populations. Data are given in mean (standard deviation). MoCA score ranges from 0 to 30, QoL items from 0 to 100.

	Present study	Reference population					
		Healthy populations				Brain tumour patients	
MoCA test		Rossetti et al. (2011)[I]	Rossetti et al. (2011)[II]	Thomann et al. (2018)[III]		Robinson et al. (2015)	Schiavolin et al. (2018)[IV]
MoCA total score[+]	24.8 (4.2)	23.7 (3.8)	25.6 (2.9)	26.1 (2.5)		26.5 (2.1)	23.2 (3.7)
EORTC-QLQ-C30		Scott et al. (2008)	Schwarz and Hinz (2001)[III]	Michelson et al. (2000)[VI]	Michelson et al. (2000)[VII]	Renovanz et al. (2018)	Scott et al. (2008)[V]
Cognitive function[+]	68.9 (26.6)	86.1 (20.0)	91.2 (17.0)	88.5 (17.7)	88.0 (17.0)	58.1 (29.8)	72.8 (26.1)
Global health status[+]	58.3 (22.2)	71.2 (22.4)	70.8 (22.1)	74.7 (22.2)	78.1 (21.3)	43.6 (22.4)	61.6 (22.2)
Social function[+]	61.7 (32.5)	87.5 (22.9)	91.0 (19.4)	90.4 (19.6)	91.3 (19.3)	50.4 (31.0)	66.8 (31.0)
Physical function[+]	75.4 (23.8)	89.8 (16.2)	90.1 (16.7)	88.0 (17.7)	91.5 (16.4)	64.7 (32.6)	
Emotional function[+]	61.2 (25.4)	76.3 (22.8)	78.7 (21.0)	78.3 (21.9)	84.0 (19.8)	46.1 (27.8)	70.9 (23.4)
Role function[+]	64.7 (34.3)	84.7 (25.4)	88.0 (22.9)	86.0 (24.4)	87.2 (24.4)		58.1 (32.9)
Fatigue[−]	43.0 (29.6)	24.1 (24.0)	17.1 (22.0)	23.4 (22.4)	18.7 (20.7)	49.7 (32.2)	37.5 (26.2)
Insomnia[−]	27.9 (33.4)	21.8 (29.7)	16.4 (27.2)	20.3 (27.5)	14.0 (24.5)		21.7 (29.2)
Pain[−]	27.0 (29.2)	20.9 (27.6)	15.4 (24.4)	20.6 (26.9)	16.0 (23.3)	33.7 (34.2)	16.5 (23.9)
EORTC-QLQ-BN20							Taphoorn et al. (2010)[VIII]
Communication deficit[−]	20.4 (26.3)						17.5 (24.5)
Visual disorder[−]	21.1 (23.3)						12.8 (18.8)
Future uncertainty[−]	44.7 (26.2)					56.2 (32.4)	37.0 (27.1)
Headache[−]	36.6 (32.6)					40.0 (35.8)	21.2 (26.2)
Seizures[−]	2.7 (14.0)						6.1 (18.5)
Hair loss[−]	8.8 (19.4)						9.4 (22.6)

[+]high score: high functionality, [−]high score: high symptomatology; [I] general population; [II] white population; [III] German healthy population; [IV] after surgery; [V] brain tumour patients; [VI] female population; [VII] male population; [VIII] baseline values

D Dose comparison

Table D.1: Dose comparison for organs at risk for the photon (XRT) and proton treatment plans (PBT) as well as the difference of both. Median, mean and standard deviation are given for each modality and patient-individual differences. P-values are derived from Wilcoxon test. Median relative changes $\Delta V_{\text{rel}} = (V_{\text{XRT}} - V_{\text{PBT}})/V_{\text{XRT}}$ and $\Delta D_{\text{rel}} = (D_{\text{XRT}} - D_{\text{PBT}})/D_{\text{XRT}}$ in per cent are given.

			PBT		XRT		XRT-PBT			
Organ at risk	**Parameter**		**Median**	**(Mean ± SD)**	**Median**	**(Mean ± SD)**	**Median**	**(Mean ± SD)**	**p-value**	ΔV_{rel}, ΔD_{rel}
Brain-CTV	$V10Gy(RBE)$	[%]	21.6	(23.4 ± 13.5)	51.3	(48.9 ± 24.1)	23.8	(25.6 ± 18.7)	< 0.001	54 %
	$V20Gy(RBE)$	[%]	13.5	(15.9 ± 11.0)	30.1	(30.9 ± 19.5)	11.1	(15.0 ± 12.4)	< 0.001	47 %
	$V30Gy(RBE)$	[%]	9.8	(11.9 ± 9.1)	17.4	(18.7 ± 12.9)	4.6	(6.9 ± 6.7)	< 0.001	37 %
	$V40Gy(RBE)$	[%]	7.5	(8.8 ± 7.6)	10.0	(11.5 ± 8.9)	1.4	(2.7 ± 3.6)	< 0.001	26 %
	$V50Gy(RBE)$	[%]	4.5	(5.9 ± 5.4)	5.8	(7.3 ± 6.1)	0.9	(1.3 ± 2.0)	< 0.001	22 %
	$V60Gy(RBE)$	[%]	0.0	(0.9 ± 1.4)	0.0	(1.0 ± 1.7)	0.0	(0.1 ± 0.9)	0.28	3 %
	D_{mean}	[Gy(RBE)]	7.3	(8.2 ± 5.2)	16.1	(15.7 ± 8.0)	7.2	(7.5 ± 4.8)	< 0.001	52 %
	$D2\%$	[Gy(RBE)]	53.8	(50.6 ± 11.2)	53.9	(51.9 ± 9.6)	0.4	(1.2 ± 3.4)	< 0.001	1 %
	$D50\%$	[Gy(RBE)]	0.0	(1.4 ± 4.5)	10.7	(10.8 ± 8.7)	9.1	(9.5 ± 7.9)	< 0.001	99 %
Brain stem	$D2\%$	[Gy(RBE)]	31.5	(30.7 ± 22.8)	44.6	(39.2 ± 15.9)	3.8	(8.4 ± 11.2)	< 0.001	15 %
Cerebellum	$V10Gy(RBE)$	[%]	1.4	(11.4 ± 19.0)	36.7	(43.6 ± 36.8)	24.7	(32.2 ± 28.1)	< 0.001	93 %
	$V20Gy(RBE)$	[%]	0.5	(8.0 ± 14.9)	10.3	(22.7 ± 27.0)	8.0	(14.7 ± 17.7)	< 0.001	87 %
	$V30Gy(RBE)$	[%]	0.0	(5.5 ± 11.3)	2.1	(10.8 ± 17.6)	0.8	(5.3 ± 8.5)	< 0.001	66 %
	$V40Gy(RBE)$	[%]	0.0	(3.8 ± 8.0)	0.0	(5.9 ± 11.8)	0.0	(2.1 ± 4.7)	< 0.001	52 %
	$V50Gy(RBE)$	[%]	0.0	(2.3 ± 5.2)	0.0	(3.3 ± 7.5)	0.0	(1.1 ± 2.8)	< 0.001	49 %
	$V60Gy(RBE)$	[%]	0.0	(0.1 ± 0.7)	0.0	(0.1 ± 0.7)	0.0	(0.0 ± 0.3)	0.89	−18 %
	D_{mean}	[Gy(RBE)]	0.4	(4.0 ± 6.8)	9.5	(12.6 ± 10.2)	7.0	(8.6 ± 5.9)	< 0.001	95 %
	$D2\%$	[Gy(RBE)]	5.6	(20.0 ± 23.0)	30.2	(31.1 ± 18.5)	9.3	(11.1 ± 10.3)	< 0.001	82 %
	$D50\%$	[Gy(RBE)]	0.0	(1.5 ± 5.3)	7.8	(10.7 ± 10.1)	6.9	(9.2 ± 8.0)	< 0.001	100 %
Cerebellum anterior	$V10Gy(RBE)$	[%]	2.5	(19.5 ± 30.1)	90.2	(63.7 ± 39.6)	35.7	(44.2 ± 36.4)	< 0.001	93 %
	$V20Gy(RBE)$	[%]	0.4	(14.5 ± 25.3)	30.9	(40.0 ± 38.7)	19.9	(25.6 ± 28.2)	< 0.001	89 %
	$V30Gy(RBE)$	[%]	0.0	(11.0 ± 21.4)	2.5	(21.7 ± 31.6)	2.0	(10.7 ± 15.8)	< 0.001	82 %
	$V40Gy(RBE)$	[%]	0.0	(8.3 ± 17.6)	0.0	(12.8 ± 23.8)	0.0	(4.5 ± 8.8)	< 0.001	50 %
	$V50Gy(RBE)$	[%]	0.0	(5.2 ± 12.0)	0.0	(7.6 ± 16.3)	0.0	(2.4 ± 6.0)	< 0.001	48 %
	$V60Gy(RBE)$	[%]	0.0	(0.3 ± 1.5)	0.0	(0.3 ± 1.3)	0.0	(0.0 ± 1.2)	0.50	−25 %
	D_{mean}	[Gy(RBE)]	0.9	(7.2 ± 12.0)	17.8	(19.2 ± 13.6)	10.2	(11.9 ± 7.6)	< 0.001	93 %
	$D2\%$	[Gy(RBE)]	12.0	(21.8 ± 23.4)	30.4	(32.8 ± 18.5)	10.3	(11.0 ± 10.7)	< 0.001	55 %
	$D50\%$	[Gy(RBE)]	0.0	(6.0 ± 13.1)	16.7	(18.4 ± 14.8)	9.9	(12.3 ± 8.9)	< 0.001	100 %
Cerebellum posterior	$V10Gy(RBE)$	[%]	0.1	(10.2 ± 18.3)	29.0	(40.8 ± 38.7)	21.2	(30.7 ± 30.2)	< 0.001	93 %
	$V20Gy(RBE)$	[%]	0.0	(7.0 ± 14.3)	5.9	(20.2 ± 26.9)	2.8	(13.1 ± 18.0)	< 0.001	89 %
	$V30Gy(RBE)$	[%]	0.0	(4.6 ± 10.6)	0.0	(9.1 ± 16.4)	0.0	(4.5 ± 7.9)	< 0.001	66 %
	$V40Gy(RBE)$	[%]	0.0	(3.0 ± 7.1)	0.0	(4.7 ± 10.5)	0.0	(1.7 ± 4.2)	< 0.001	51 %
	$V50Gy(RBE)$	[%]	0.0	(1.7 ± 4.5)	0.0	(2.6 ± 6.6)	0.0	(0.9 ± 2.4)	< 0.001	44 %
	$V60Gy(RBE)$	[%]	0.0	(0.1 ± 0.7)	0.0	(0.1 ± 0.7)	0.0	(0.0 ± 0.2)	0.61	30 %
	D_{mean}	[Gy(RBE)]	0.1	(3.4 ± 6.4)	8.6	(11.5 ± 10.1)	6.3	(8.1 ± 6.0)	< 0.001	98 %
	$D2\%$	[Gy(RBE)]	0.7	(17.4 ± 22.2)	23.3	(26.6 ± 19.8)	7.6	(9.2 ± 9.2)	< 0.001	94 %
	$D50\%$	[Gy(RBE)]	0.0	(1.3 ± 5.0)	7.4	(10.1 ± 9.6)	7.4	(8.8 ± 7.7)	< 0.001	100 %
Chiasm	$D2\%$	[Gy(RBE)]	45.1	(32.3 ± 23.9)	50.4	(37.5 ± 20.1)	2.2	(5.3 ± 8.5)	< 0.001	9 %
Cochlea contralateral	D_{mean}	[Gy(RBE)]	0.0	(1.9 ± 5.4)	5.1	(8.9 ± 9.0)	4.4	(7.1 ± 7.0)	< 0.001	100 %
	$D2\%$	[Gy(RBE)]	0.0	(2.8 ± 7.0)	7.3	(10.6 ± 10.7)	4.9	(7.8 ± 7.8)	< 0.001	100 %
Cochlea ipsilateral	D_{mean}	[Gy(RBE)]	2.4	(16.0 ± 20.3)	17.6	(21.5 ± 18.8)	3.4	(5.5 ± 8.3)	< 0.001	78 %
	$D2\%$	[Gy(RBE)]	4.9	(19.8 ± 22.9)	21.7	(24.4 ± 20.1)	2.8	(4.6 ± 8.8)	< 0.001	67 %
Frontal lobe contralateral	$V10Gy(RBE)$	[%]	4.8	(21.4 ± 28.1)	27.2	(39.6 ± 37.8)	8.4	(18.2 ± 32.4)	< 0.001	74 %
	$V20Gy(RBE)$	[%]	0.8	(12.8 ± 21.2)	9.3	(24.3 ± 31.6)	2.1	(11.4 ± 19.1)	< 0.001	56 %
	$V30Gy(RBE)$	[%]	0.2	(9.8 ± 18.1)	3.3	(14.6 ± 24.2)	0.3	(4.8 ± 10.1)	< 0.001	55 %
	$V40Gy(RBE)$	[%]	0.0	(7.3 ± 14.9)	0.4	(9.1 ± 18.3)	0.0	(1.8 ± 5.4)	< 0.001	31 %
	$V50Gy(RBE)$	[%]	0.0	(5.1 ± 12.0)	0.0	(6.0 ± 13.8)	0.0	(0.9 ± 3.3)	< 0.001	16 %
	$V60Gy(RBE)$	[%]	0.0	(1.3 ± 4.2)	0.0	(1.7 ± 5.7)	0.0	(0.4 ± 2.9)	0.16	25 %
	D_{mean}	[Gy(RBE)]	1.6	(7.2 ± 10.7)	8.4	(13.0 ± 13.2)	3.6	(5.8 ± 7.3)	< 0.001	78 %
	$D2\%$	[Gy(RBE)]	14.1	(24.6 ± 24.7)	33.0	(32.2 ± 20.5)	4.1	(7.6 ± 10.5)	< 0.001	37 %
	$D50\%$	[Gy(RBE)]	0.0	(4.8 ± 10.8)	6.9	(11.3 ± 13.7)	3.4	(6.5 ± 10.5)	< 0.001	100 %

Abbreviation: SD, standard deviation.

Table D.1 [continued]: Dose comparison for organs at risk for the photon (XRT) and proton treatment plans (PBT) as well as the difference of both. Median, mean and standard deviation are given for each modality and patient-individual differences. P-values are derived from Wilcoxon test. Median relative changes $\Delta V_{\text{rel}} = (V_{\text{XRT}} - V_{\text{PBT}})/V_{\text{XRT}}$ and $\Delta D_{\text{rel}} = (D_{\text{XRT}} - D_{\text{PBT}})/D_{\text{XRT}}$ in per cent are given.

			PBT		XRT		XRT-PBT			
Organ at risk	**Parameter**		**Median**	**(Mean ± SD)**	**Median**	**(Mean ± SD)**	**Median**	**(Mean ± SD)**	**p-value**	ΔV_{rel}, ΔD_{rel}
Frontal lobe	$V10Gy(RBE)$	[%]	29.3	(34.7 ± 31.7)	33.9	(43.7 ± 35.7)	6.5	(9.0 ± 32.1)	0.017	30 %
ipsilateral	$V20Gy(RBE)$	[%]	12.5	(21.7 ± 24.7)	15.6	(31.1 ± 32.2)	2.5	(9.4 ± 19.0)	< 0.001	35 %
	$V30Gy(RBE)$	[%]	7.6	(16.0 ± 20.9)	9.1	(21.0 ± 24.6)	1.2	(5.0 ± 9.1)	< 0.001	23 %
	$V40Gy(RBE)$	[%]	3.5	(11.9 ± 18.1)	4.3	(13.3 ± 18.6)	0.0	(1.4 ± 4.5)	0.007	19 %
	$V50Gy(RBE)$	[%]	1.0	(8.6 ± 15.6)	2.0	(9.4 ± 15.5)	0.0	(0.8 ± 2.2)	0.011	13 %
	$V60Gy(RBE)$	[%]	0.0	(2.4 ± 6.2)	0.0	(3.3 ± 9.2)	0.0	(0.9 ± 4.0)	0.53	9 %
	D_{mean}	[Gy(RBE)]	7.6	(11.6 ± 12.3)	10.1	(15.5 ± 13.6)	2.3	(3.8 ± 7.3)	< 0.001	40 %
	$D2\%$	[Gy(RBE)]	45.8	(35.8 ± 23.5)	49.8	(40.4 ± 20.0)	0.8	(4.7 ± 7.8)	< 0.001	2 %
	$D50\%$	[Gy(RBE)]	0.6	(9.4 ± 15.1)	4.9	(13.6 ± 15.9)	1.3	(4.2 ± 10.8)	< 0.001	93 %
Hippocampus	D_{mean}	[Gy(RBE)]	0.1	(5.3 ± 10.7)	12.8	(16.0 ± 12.0)	9.5	(10.7 ± 8.3)	< 0.001	99 %
contralateral	$D2\%$	[Gy(RBE)]	0.6	(13.0 ± 19.1)	22.6	(26.2 ± 16.1)	11.4	(13.2 ± 11.4)	< 0.001	95 %
	$D50\%$	[Gy(RBE)]	0.0	(4.3 ± 10.3)	13.1	(15.5 ± 12.1)	9.5	(11.1 ± 8.8)	< 0.001	100 %
Hippocampus	D_{mean}	[Gy(RBE)]	12.2	(18.5 ± 19.1)	28.8	(28.2 ± 16.8)	8.3	(9.7 ± 8.6)	< 0.001	48 %
ipsilateral	$D2\%$	[Gy(RBE)]	39.6	(33.1 ± 23.2)	44.4	(40.1 ± 18.0)	2.9	(7.0 ± 11.1)	< 0.001	6 %
	$D50\%$	[Gy(RBE)]	6.5	(17.3 ± 20.9)	26.6	(28.0 ± 17.8)	9.1	(10.7 ± 10.0)	< 0.001	67 %
Hippocampi	D_{mean}	[Gy(RBE)]	5.9	(11.0 ± 12.2)	19.8	(21.2 ± 12.7)	8.7	(10.2 ± 7.1)	< 0.001	54 %
bilateral	$D2\%$	[Gy(RBE)]	36.3	(31.9 ± 23.2)	42.0	(39.3 ± 18.0)	2.9	(7.4 ± 11.0)	< 0.001	6 %
Lacrimal gland	D_{mean}	[Gy(RBE)]	0.0	(0.6 ± 2.5)	4.4	(5.6 ± 5.6)	4.1	(5.1 ± 5.1)	< 0.001	100 %
contralateral	$D2\%$	[Gy(RBE)]	0.0	(1.4 ± 5.5)	5.6	(7.7 ± 7.4)	5.1	(6.3 ± 6.4)	< 0.001	100 %
Lacrimal gland	D_{mean}	[Gy(RBE)]	0.4	(7.9 ± 12.6)	9.4	(13.6 ± 12.5)	4.2	(5.7 ± 8.9)	< 0.001	94 %
ipsilateral	$D2\%$	[Gy(RBE)]	2.4	(13.6 ± 18.3)	12.8	(18.8 ± 16.0)	5.4	(5.2 ± 10.9)	< 0.001	84 %
Lens	D_{mean}	[Gy(RBE)]	0.0	(0.3 ± 1.1)	3.4	(4.7 ± 5.2)	3.3	(4.4 ± 4.7)	< 0.001	100 %
contralateral	$D2\%$	[Gy(RBE)]	0.0	(0.6 ± 1.6)	4.2	(5.5 ± 6.1)	3.9	(5.0 ± 5.4)	< 0.001	100 %
Lens	D_{mean}	[Gy(RBE)]	0.0	(2.2 ± 6.0)	4.4	(9.6 ± 10.0)	4.1	(7.4 ± 8.7)	< 0.001	100 %
ipsilateral	$D2\%$	[Gy(RBE)]	0.0	(3.2 ± 7.8)	4.8	(10.8 ± 11.2)	4.3	(7.6 ± 9.8)	< 0.001	100 %
Optic nerve ipsilateral	$D2\%$	[Gy(RBE)]	4.1	(17.1 ± 20.4)	24.2	(24.9 ± 18.5)	4.7	(7.9 ± 10.3)	< 0.001	82 %
Optic nerve contralateral	$D2\%$	[Gy(RBE)]	40.6	(30.1 ± 24.4)	47.2	(34.4 ± 21.7)	1.9	(4.3 ± 8.9)	< 0.001	13 %
Pituitary	D_{mean}	[Gy(RBE)]	4.7	(18.7 ± 22.0)	28.7	(27.7 ± 20.2)	5.1	(9.0 ± 10.4)	< 0.001	81 %
	$D2\%$	[Gy(RBE)]	12.3	(23.4 ± 23.7)	36.1	(31.9 ± 21.1)	4.9	(8.5 ± 10.5)	< 0.001	48 %
Skin	$V10Gy(RBE)$	[cm^3]	40.5	(42.1 ± 29.0)	82.4	(83.4 ± 49.6)	32.3	(41.4 ± 30.9)	< 0.001	51 %
	$V20Gy(RBE)$	[cm^3]	29.7	(27.5 ± 27.7)	24.5	(36.4 ± 34.8)	4.0	(8.9 ± 14.9)	< 0.001	36 %
	$V30Gy(RBE)$	[cm^3]	8.4	(17.4 ± 22.0)	9.0	(18.6 ± 22.0)	0.0	(1.2 ± 8.1)	0.12	10 %
	$V40Gy(RBE)$	[cm^3]	2.8	(10.9 ± 15.9)	1.9	(10.2 ± 14.1)	0.0	(-0.7 ± 6.3)	0.46	−2 %
	$V50Gy(RBE)$	[cm^3]	0.0	(4.8 ± 8.1)	0.0	(3.9 ± 7.3)	0.0	(-0.9 ± 4.5)	0.088	−12 %
	$V60Gy(RBE)$	[cm^3]	0.0	(0.2 ± 1.3)	0.0	(0.2 ± 1.9)	0.0	(0.0 ± 0.7)	0.051	−46 %
	$D2\%$	[Gy(RBE)]	30.1	(32.0 ± 17.0)	30.8	(32.3 ± 15.3)	0.6	(0.3 ± 5.3)	0.36	2 %
	$D5\%$	[Gy(RBE)]	23.4	(24.0 ± 16.2)	21.7	(25.8 ± 13.8)	2.1	(1.8 ± 5.6)	0.002	9 %
	$D15\%$	[Gy(RBE)]	0.5	(4.8 ± 9.2)	13.0	(14.1 ± 8.6)	8.4	(9.3 ± 6.7)	< 0.001	95 %
	$D25\%$	[Gy(RBE)]	0.0	(1.1 ± 4.3)	6.5	(8.3 ± 5.8)	5.8	(7.2 ± 5.4)	< 0.001	99 %
	$D35\%$	[Gy(RBE)]	0.0	(0.4 ± 1.7)	4.1	(4.9 ± 4.0)	4.1	(4.5 ± 3.6)	< 0.001	99 %
	$D45\%$	[Gy(RBE)]	0.0	(0.1 ± 0.4)	2.4	(2.8 ± 2.4)	2.2	(2.7 ± 2.2)	< 0.001	99 %
	$D55\%$	[Gy(RBE)]	0.0	(0.0 ± 0.1)	1.2	(1.7 ± 1.6)	1.2	(1.6 ± 1.5)	< 0.001	100 %
Temporal lobe	$V10Gy(RBE)$	[%]	0.2	(12.5 ± 21.8)	37.4	(43.4 ± 34.8)	23.5	(31.0 ± 32.7)	< 0.001	99 %
contralateral	$V20Gy(RBE)$	[%]	0.0	(5.0 ± 12.4)	5.3	(20.7 ± 27.7)	3.8	(15.7 ± 22.8)	< 0.001	96 %
	$V30Gy(RBE)$	[%]	0.0	(2.7 ± 7.1)	0.1	(7.1 ± 14.4)	0.1	(4.4 ± 8.6)	< 0.001	78 %
	$V40Gy(RBE)$	[%]	0.0	(1.8 ± 5.4)	0.0	(2.8 ± 7.8)	0.0	(1.0 ± 2.9)	< 0.001	47 %
	$V50Gy(RBE)$	[%]	0.0	(1.0 ± 3.4)	0.0	(1.4 ± 4.6)	0.0	(0.4 ± 1.6)	< 0.001	37 %
	$V60Gy(RBE)$	[%]	0.0	(0.0 ± 0.0)	0.0	(0.0 ± 0.1)	0.0	(0.0 ± 0.1)	0.18	86 %
	D_{mean}	[Gy(RBE)]	0.2	(3.1 ± 5.7)	9.4	(11.8 ± 8.5)	7.6	(8.7 ± 6.4)	< 0.001	97 %
	$D2\%$	[Gy(RBE)]	1.7	(14.6 ± 19.5)	23.1	(27.0 ± 15.8)	10.5	(12.4 ± 10.7)	< 0.001	86 %
	$D50\%$	[Gy(RBE)]	0.0	(2.1 ± 5.3)	8.3	(10.7 ± 8.5)	7.5	(8.6 ± 7.3)	< 0.001	100 %
Temporal lobe	$V10Gy(RBE)$	[%]	42.4	(45.3 ± 29.1)	74.6	(69.9 ± 29.8)	21.6	(24.6 ± 22.5)	< 0.001	38 %
ipsilateral	$V20Gy(RBE)$	[%]	26.5	(33.7 ± 28.8)	51.5	(51.3 ± 31.4)	14.2	(17.6 ± 16.6)	< 0.001	38 %
	$V30Gy(RBE)$	[%]	18.5	(26.6 ± 27.2)	29.7	(36.4 ± 30.1)	8.2	(9.8 ± 10.4)	< 0.001	38 %
	$V40Gy(RBE)$	[%]	12.0	(21.2 ± 25.7)	15.3	(25.6 ± 27.5)	2.9	(4.4 ± 6.3)	< 0.001	24 %
	$V50Gy(RBE)$	[%]	7.1	(15.8 ± 22.7)	7.8	(18.1 ± 23.8)	1.0	(2.3 ± 3.9)	< 0.001	18 %
	$V60Gy(RBE)$	[%]	0.0	(3.0 ± 9.4)	0.0	(3.3 ± 9.7)	0.0	(0.3 ± 2.2)	0.37	14 %
	D_{mean}	[Gy(RBE)]	12.6	(17.2 ± 14.4)	23.3	(24.9 ± 14.0)	7.5	(7.7 ± 5.4)	< 0.001	34 %
	$D2\%$	[Gy(RBE)]	53.7	(44.6 ± 17.9)	54.1	(48.6 ± 13.0)	0.6	(4.0 ± 7.5)	< 0.001	1 %
	$D50\%$	[Gy(RBE)]	3.5	(13.8 ± 18.4)	20.8	(23.5 ± 16.4)	8.3	(9.6 ± 8.2)	< 0.001	75 %

Abbreviation: SD, standard deviation.

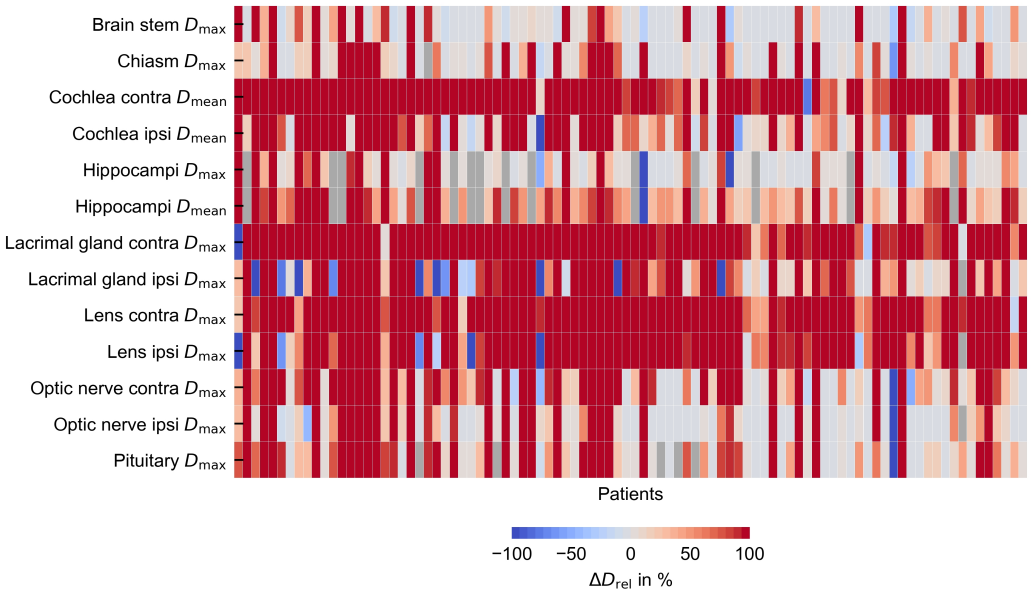


Figure D.1: Relative difference of selected DVH parameters between photon and proton treatment plans for each patient. The difference between the XRT and PBT value is normalised to the XRT value.

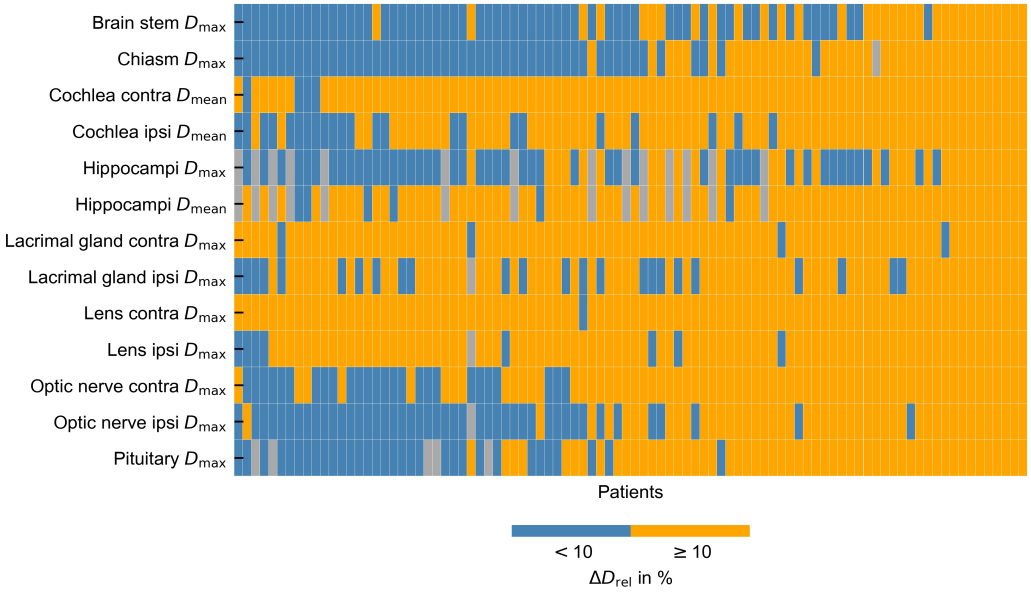


Figure D.2: Relative difference of selected DVH parameters between photon and proton treatment plans exceeding a threshold of 10 % for each patient. The difference between the XRT and PBT value is normalised to the XRT value.

Table D.2: Differences in the near-maximum parameter $D2\%$ and the mean dose D_{mean} for organs at risk with patient classification according to tumour location. Absolute differences $\Delta D_{abs} = D_{XRT} - D_{PBT}$ are given in Gy(RBE) and relative differences are given $\Delta D_{rel} = (D_{XRT} - D_{PBT})/D_{XRT}$ in per cent. The median values of each dosimetric parameter for each localisation group are given. P-values are derived from Wilcoxon test.

	Frontal lobe				Parietal lobe				Temporal lobe				Multiple lobes			
	N	ΔD_{abs} in Gy(RBE)	ΔD_{rel} in %	p-value	N	ΔD_{abs} in Gy(RBE)	ΔD_{rel} in %	p-value	N	ΔD_{abs} in Gy(RBE)	ΔD_{rel} in %	p-value	N	ΔD_{abs} in Gy(RBE)	ΔD_{rel} in %	p-value
***D*2%**																
Brain-CTV	16	0.1	0	0.38	5	0.6	1	0.043	25	0.8	2	0.035	31	0.3	0	0.017
Brain stem	16	15.7	90	0.002	5	9.1	70	0.080	25	0.8	2	0.045	31	2.0	6	0.006
Cerebellum	16	10.2	100	< 0.001	5	17.2	83	0.043	25	9.0	29	0.001	31	9.6	51	< 0.001
Cerebellum anterior	16	12.2	100	< 0.001	5	16.8	53	0.043	25	6.8	34	0.002	31	9.7	43	< 0.001
Cerebellum posterior	16	6.6	100	< 0.001	5	3.9	98	0.043	25	5.6	48	0.001	31	9.4	87	< 0.001
Chiasm	15	3.9	16	0.036	5	7.3	100	0.042	25	0.3	1	0.093	31	3.0	5	0.003
Cochlea contra	16	1.8	100	0.001	5	1.6	100	0.025	25	7.2	100	< 0.001	31	7.8	100	< 0.001
Cochlea ipsi	16	3.5	100	0.001	5	2.4	100	0.034	25	0.4	1	0.19	31	3.7	64	< 0.001
Frontal lobe contra	15	0.2	0	0.19	5	8.7	100	0.039	25	5.7	76	0.002	31	2.7	6	< 0.001
Frontal lobe ipsi	5	0.1	0	0.69	4	5.9	55	0.068	23	0.8	2	0.021	25	0.4	1	0.013
Hippocampus contra	16	9.9	100	0.001	4	22.1	100	0.066	23	15.8	83	< 0.001	27	14.1	97	< 0.001
Hippocampus ipsi	15	14.0	93	0.001	4	1.2	2	0.14	19	1.5	3	0.036	26	0.8	1	0.007
Hippocampi	15	15.6	94	0.001	4	1.7	3	0.068	19	1.2	2	0.044	26	0.9	2	0.014
Lacrimal gland contra	16	5.7	99	< 0.001	5	7.6	100	0.025	25	3.6	100	< 0.001	31	5.5	100	< 0.001
Lacrimal gland ipsi	15	1.0	26	0.91	5	5.6	100	0.034	25	6.5	64	0.034	31	5.8	70	< 0.001
Lens contra	16	3.2	97	< 0.001	5	1.5	100	0.025	25	4.3	100	< 0.001	31	4.1	100	< 0.001
Lens ipsi	15	3.9	94	0.001	5	1.8	100	0.025	25	5.7	100	< 0.001	31	4.7	99	< 0.001
Optic nerve contra	16	4.6	80	0.017	5	2.8	100	0.025	25	3.6	30	0.001	31	8.3	67	< 0.001
Optic nerve ipsi	15	1.0	5	0.22	5	3.0	100	0.034	25	-0.1	0	0.21	31	3.2	13	0.002
Pituitary	16	6.7	53	0.004	5	5.3	100	0.039	23	1.1	2	0.003	30	7.7	46	< 0.001
Skin	16	-2.8	-5	0.35	5	5.4	15	0.080	25	0.8	4	0.76	31	0.1	0	0.98
Temporal lobe contra	16	12.9	98	0.001	5	30.2	100	0.039	25	9.9	51	< 0.001	31	11.6	98	< 0.001
Temporal lobe ipsi	16	6.9	15	0.007	3	0.6	1	0.11	19	0.3	1	0.077	27	0.5	1	0.031
D_{mean}																
Brain-CTV	16	9.6	58	< 0.001	5	10.5	57	0.043	25	3.2	51	< 0.001	31	8.8	47	< 0.001
Cerebellum	16	3.5	100	< 0.001	5	2.9	89	0.043	25	9.0	88	< 0.001	31	10.4	94	< 0.001
Cerebellum anterior	16	7.2	100	< 0.001	5	8.5	91	0.043	25	11.8	91	< 0.001	31	12.9	94	< 0.001
Cerebellum posterior	16	3.0	100	< 0.001	5	2.0	99	0.042	25	7.9	92	< 0.001	31	10.1	98	< 0.001
Cochlea contra	16	1.5	100	< 0.001	5	1.2	100	0.025	25	6.7	100	< 0.001	31	5.9	100	< 0.001
Cochlea ipsi	16	3.2	100	< 0.001	5	1.8	100	0.034	25	1.3	5	0.016	31	4.3	84	< 0.001
Frontal lobe contra	15	12.9	37	0.047	5	15.0	100	0.039	25	1.6	88	0.073	31	8.6	69	< 0.001
Frontal lobe ipsi	5	2.4	25	0.50	4	12.4	89	0.068	23	0.5	34	0.58	25	7.7	23	0.004
Hippocampus contra	16	7.3	100	< 0.001	4	16.8	100	0.066	23	9.8	97	< 0.001	27	8.9	99	< 0.001
Hippocampus ipsi	15	9.9	98	0.001	4	6.2	27	0.068	19	5.8	18	0.001	26	3.8	12	< 0.001
Hippocampi	15	8.7	99	0.001	4	11.3	51	0.068	19	7.9	46	< 0.001	26	7.5	41	< 0.001
Lacrimal gland contra	16	4.1	99	< 0.001	5	3.1	100	0.025	25	2.9	100	< 0.001	31	4.7	100	< 0.001
Lacrimal gland ipsi	15	2.4	51	0.14	5	3.5	100	0.025	25	6.0	87	0.006	31	5.0	89	< 0.001
Pituitary	16	5.5	73	0.001	5	3.9	100	0.039	23	2.1	6	0.001	30	8.2	81	< 0.001
Temporal lobe contra	16	5.7	99	0.001	5	11.6	100	0.039	25	7.4	92	< 0.001	31	9.3	100	< 0.001
Temporal lobe ipsi	16	8.8	68	< 0.001	3	7.3	36	0.11	19	5.3	26	0.001	27	6.3	23	< 0.001

Abbreviations: N, number of patients; CTV, clinical target volume; ipsi, ipsilateral; contra, contralateral.

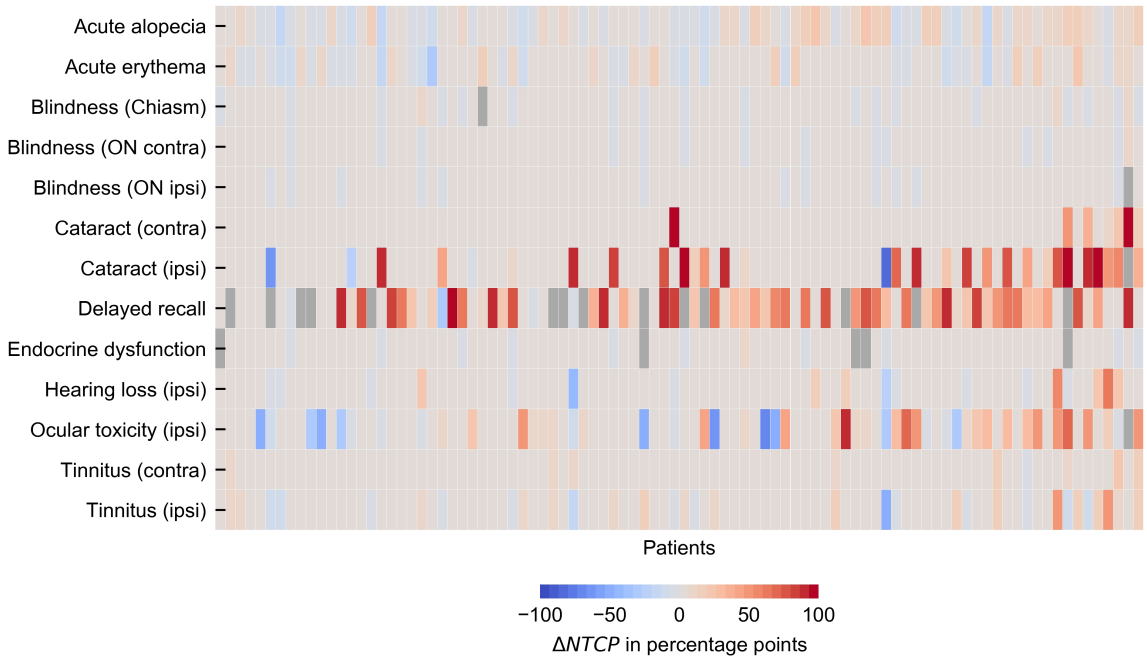


Figure D.3: Difference in NTCP between XRT and PBT for selected side effects for each patient.

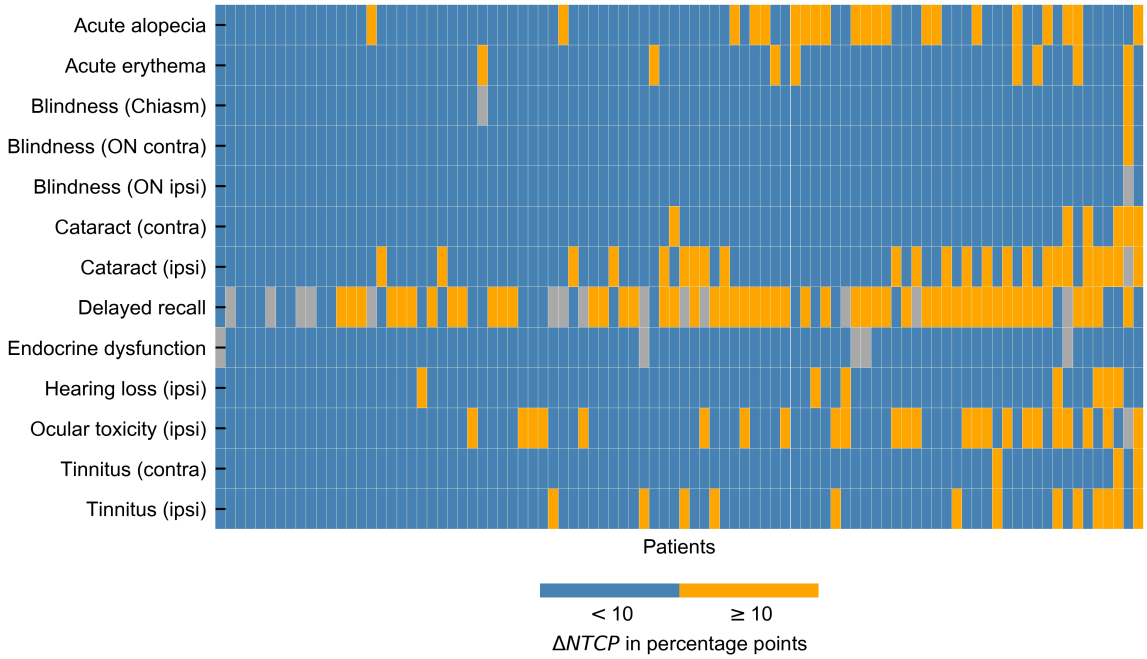


Figure D.4: Difference in NTCP between XRT and PBT for selected side effects exceeding a threshold of 10 percentage points for each patient.

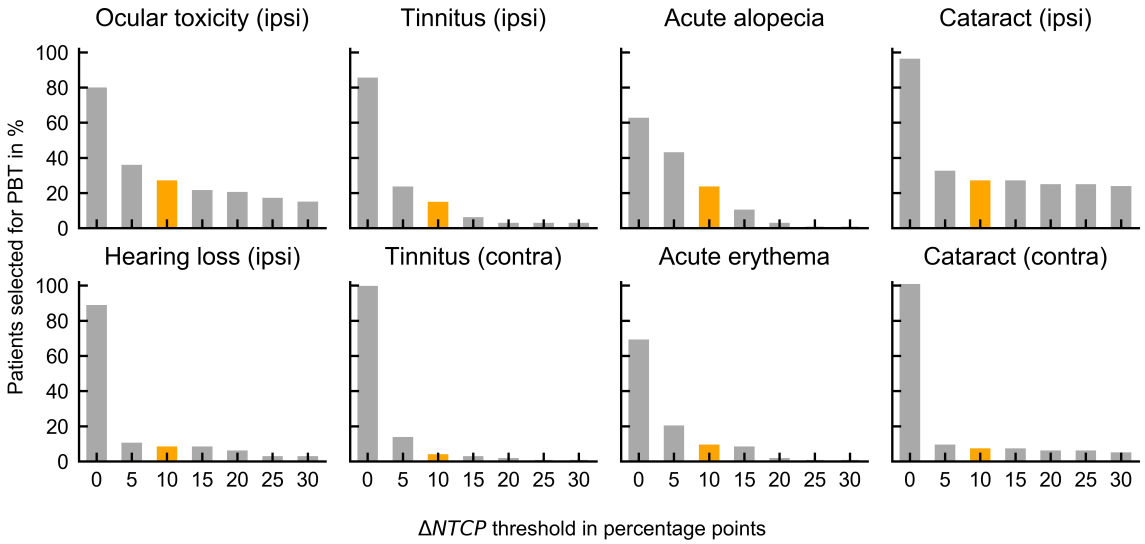


Figure D.5: Patient selection for different side effects and different thresholds. The applied threshold of 10 percentage points in this study is shaded in orange.

Table D.3: NTCP differences with patient classification according to tumour location. Tumour localisations other than those indicated are not considered. Median estimates for photon plans ($NTCP_{XRT}$) and proton plans ($NTCP_{PBT}$) in per cent as well as median difference ($\Delta NTCP = NTCP_{XRT} - NTCP_{PBT}$) in percentage points are given for each localisation group. P-values are derived from Wilcoxon test.

	Frontal lobe					Parietal lobe					Temporal lobe					Multiple lobes				
Side effect	**N**	$NTCP_{XRT}$	$NTCP_{PBT}$	$\Delta NTCP$	**p-value**	**N**	$NTCP_{XRT}$	$NTCP_{PBT}$	$\Delta NTCP$	**p-value**	**N**	$NTCP_{XRT}$	$\Delta NTCP$	$\Delta NTCP$	**p-value**	**N**	$NTCP_{XRT}$	$NTCP_{PBT}$	$\Delta NTCP$	**p-value**
Alopecia	16	78	79	-1.5	0.76	5	80	69	11.2	0.043	25	50	47	1.7	0.11	31	79	77	0.2	0.33
Blindness (chiasm)	15	0	0	0.0	0.017	5	0	0	0.0	0.043	25	1	0	0.0	0.17	31	0	0	0.0	0.004
Blindness (ON contra)	16	0	0	0.0	0.030	5	0	0	0.0	0.043	25	0	0	0.0	0.069	31	0	0	0.0	0.002
Blindness (ON ipsi)	15	0	0	0.0	0.027	5	0	0	0.0	0.043	25	0	0	0.0	0.074	31	0	0	0.0	0.002
Cataract contra	16	0	0	0.1	< 0.001	5	0	0	0.0	0.043	25	0	0	0.1	< 0.001	31	0	0	0.2	< 0.001
Cataract ipsi	15	0	0	0.2	0.005	5	0	0	0.0	0.043	25	0	0	0.4	< 0.001	31	0	0	0.3	< 0.001
Delayed recall	15	61	3	25.0	0.001	4	88	21	43.9	0.068	19	81	21	45.4	< 0.001	26	91	9	10.4	< 0.001
Endocrine dysfunction	16	0	0	0.0	0.015	5	0	0	0.0	0.043	23	0	0	0.0	0.013	30	0	0	0.0	0.004
Erythema	16	36	39	-0.8	0.78	5	39	21	18.4	0.068	25	18	18	0.0	0.33	31	36	39	0.0	0.89
Hearing loss ipsi	16	0	0	0.0	0.007	5	0	0	0.0	0.043	25	0	0	0.0	0.17	31	0	0	0.0	0.004
Necrosis (brain stem)	16	0	0	0.0	0.005	5	0	0	0.0	0.043	25	0	0	0.0	< 0.001	31	0	0	0.0	< 0.001
Ocular toxicity	15	49	52	0.1	0.91	5	2	1	1.3	0.043	25	6	2	3.0	0.093	31	52	15	2.4	0.001
Tinnitus contra	16	0	0	0.1	0.007	5	0	0	0.1	0.043	25	1	0	0.6	< 0.001	31	1	0	0.4	< 0.001
Tinnitus ipsi	16	0	0	0.2	0.002	5	0	0	0.1	0.043	25	19	16	0.4	0.17	31	6	0	0.3	0.007

Abbreviations: N, number of patients; ON, optic nerve; ipsi, ipsilateral; contra, contralateral.

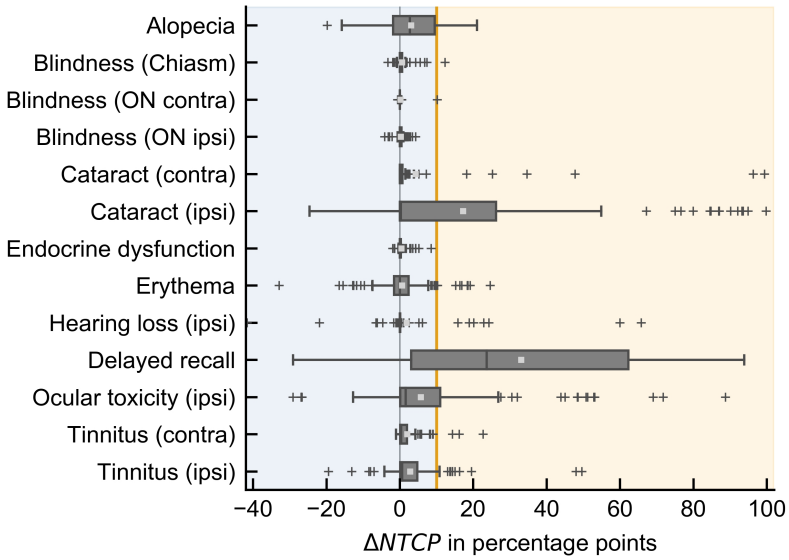


Figure D.6: NTCP differences between XRT and PBT plans for different side effects excluding patients with re-irradiation. A threshold of 10 percentage points is indicated.

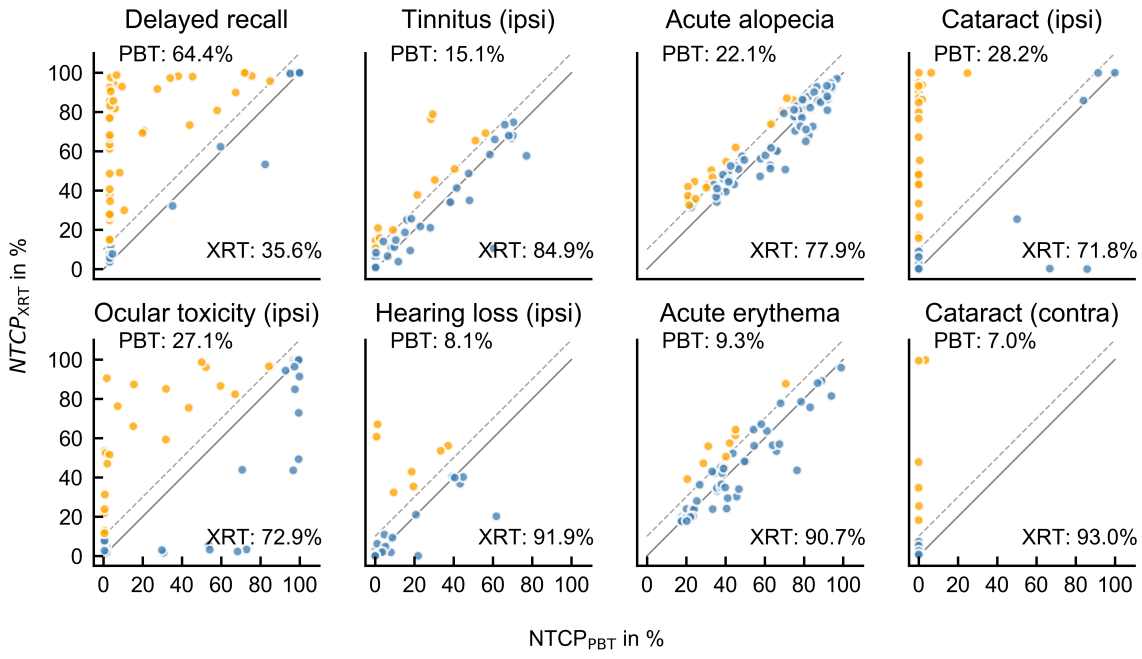


Figure D.7: NTCP values for XRT and PBT plans for selected side effects excluding patients with re-irradiation. The solid line represents equal NTCP values for both treatment modalities. A threshold of 10 percentage points is indicated as dashed line. The percentage of patients selected for each treatment is given.

Danksagung

Ich danke Professor Dr. Steffen Löck für die Bereitstellung meines Promotionsthemas, die fachliche Unterstützung, die gute Zusammenarbeit und die kritischen Nachfragen sowie für die Übernahme des Erstgutachtens.

Professor Dr. Johannes A. Langendijk danke ich für die Bereitschaft, meine Dissertation zu begutachten.

Mein Dank gebührt Professor Dr. Armin Lühr für die vielen Denkanstöße, spannenden Diskussionen und die stetige Begleitung sowie das Korrekturlesen der Arbeit.

Professorin Dr. Mechthild Krause gilt mein Dank für die Bereitstellung des klinischen Bearbeitungsrahmens sowie für das Korrekturlesen der finalen Version der Arbeit.

Für die Bereitstellung und den Transfer der Patientendaten und Bestrahlungspläne sowie für die gute Zusammenarbeit bedanke ich mich bei allen Kollegen der beiden externen Protonentherapiezentren. Das sind am Westdeutschen Protonentherapiezentrum in Essen: Professorin Dr. Beate Zimmermann, Dr. Dirk Geismar, Xavier Vermeren sowie Dr. Sabine Frisch. Am Massachusets General Hospital in Boston möchte ich mich bei Professorin Dr. Helen A. Shih, Marc Bussière, Meghan Bussière, Dr. Emily F. Schapira, Jillian E. Daly sowie Nayan Lamba.

Ich danke Professorin Dr. Esther G. C. Troost für die klinische Abnahme der Bestrahlungspläne im Rahmen der Planvergleichsstudie.

Ich danke Professor Dr. Michael Baumann für sein Engagement für die Protonentherapiestudien und seine motivierenden und hilfreichen Kommentare zu den Manuskripten.

Für die Diskussionen zu medizinischen Themen sowie das Korrekturlesen der medizinischen Abschnitte meiner Arbeit danke ich Linda Agolli und Dr. Rebecca Bütof. Ihnen und auch Chiara Valentini danke ich für die Konturierung unzähliger Hirnstrukturen. Dr. Linda Hein unterstützte mich bei der Datenakquise des MoCA-Tests sowie beim Monitoring der Studiendaten. Ich möchte Monique Falk und Daniel Büttner für ihre Mithilfe bei der Ausarbeitung der Ethikanträge danken.

Ich danke allen Ärzten, die an der Interobserver-Studie teilgenommen haben.

Für das sorgfältige Korrekturlesen einzelner Abschnitte meiner Arbeit danke ich Maria Tschiche. Ich danke den Mitgliedern der Arbeitsgruppe Modellierung und Biostatistik in der Radioonkologie für die angenehme Arbeitsatmosphäre. Meinen Kollegen und Freunden bin ich für den regen Austausch während der schönen Kaffeerunden dankbar.

Mein besonderer Dank gilt meiner Frau Anna-Louisa Biehl für ihre Unterstützung beim Schreiben und Korrekturlesen der Arbeit, für die vielen Diskussionen und Ermutigungen und dafür, dass sie mich an die wirklich wichtigen Dinge erinnert hat.

Erklärungen

Technische Universität Dresden
Medizinische Fakultät Carl Gustav Carus
Promotionsordnung vom 24.10.2014

Erklärungen zur Eröffnung des Promotionsverfahrens

1. Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

2. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten: Prof. Dr. Steffen Löck, Prof. Dr. Armin Lühr, Prof. Dr. Mechthild Krause, Prof. Dr. Esther G. C. Troost, Linda Agolli, Dr. Rebecca Bütof, Chiara Valentini.

3. Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

4. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

5. Die Inhalte dieser Dissertation wurden in folgender Form veröffentlicht:

    - Dutz A, Agolli L, Bütof R, Valentinti C, Baumann M, Lühr A, Löck S, Krause M. 2020. Neurocognitive function and quality of life after proton beam therapy for brain tumour patients. Radiother Oncol [E-Pub ahead of print] DOI: 10.1016/j.radonc.2019.12.024.
    - Dutz A, Lühr A, Agolli L, Troost EGC, Krause M, Baumann M, Vermeren X, Geismar D, Schapira EF, Bussière M, Daly JE, Bussière MR, Timmermann B, Shih HA, Löck S. 2019. Development and validation of NTCP models for acute side-effects resulting from proton beam therapy of brain tumours. Radiother Oncol 130:164-171.
    - Dutz A, Agolli L, Valentinti C, Bütof R, Troost EGC, Baumann M, Lühr A, Krause M, Löck S. 2019. PV-0361: Minor changes in neurocognition and quality of life after proton therapy for brain tumour patients. Radiother and Oncol 133:S178-S179.

- Dutz A, Agolli L, Troost EGC, Krause M, Baumann M, Lühr A, Vermeren X, Geismar D, Timmermann B, Löck S. 2018. OC-0511: NTCP modelling and external validation of early side effects for proton therapy of brain tumours. Radiother Oncol 127:S266-S267.
- Dutz A, Agolli L, Troost EGC, Krause M, Baumann M, Lühr A, Löck S. 2017. V134: Modelling of NTCP for acute side effects in patients with prostate cancer or brain tumours receiving proton therapy. Biomed Tech (Berl) 62:S304.
- Dutz A, Agolli L, Troost EGC, Krause M, Baumann M, Lühr A, Löck S. 2017. EP-1595: Development of NTCP models for patients with prostate cancer or brain tumours receiving proton therapy. Radiother Oncol 123:S860.

6. Ich bestätige, dass es keine zurückliegenden erfolglosen Promotionsverfahren gab.
7. Ich bestätige, dass ich die Promotionsordnung der Medizinischen Fakultät Carl Gustav Carus der Technischen Universität Dresden anerkenne.
8. Ich habe die Zitierrichtlinien für Dissertationen an der Medizinischen Fakultät der Technischen Universität Dresden zur Kenntnis genommen und befolgt.
9. Ich bin mit den "Richtlinien zur Sicherung guter wissenschaftlicher Praxis, zur Vermeidung wissenschaftlichen Fehlverhaltens und für den Umgang mit Verstößen" der Technischen Universität Dresden einverstanden.

Dresden, 28. August 2020

Erklärung über die Einhaltung gesetzlicher Bestimmungen

Hiermit bestätige ich die Einhaltung der folgenden aktuellen gesetzlichen Vorgaben im Rahmen meiner Dissertation

- ☒ das zustimmende Votum der Ethikkommission bei Klinischen Studien, epidemiologischen Untersuchungen mit Personenbezug oder Sachverhalten, die das Medizinproduktegesetz betreffen
  Aktenzeichen: EK219062016, EK48012019, EK566122019
- ☐ die Einhaltung der Bestimmungen des Tierschutzgesetzes
- ☐ die Einhaltung des Gentechnikgesetzes
- ☒ die Einhaltung von Datenschutzbestimmungen der Medizinischen Fakultät und des Universitätsklinikums Carl Gustav Carus.

Dresden, 28. August 2020