PhD thesis

Patrik Brodin

Risk-based optimization of photon and proton radiotherapy for pediatric medulloblastoma

Academic advisors:  
- Associate professor Per Munck af Rosenschöld, Ph.D.
- Associate professor Thomas Björk-Eriksson, M.D., Ph.D.
- Medical physicist Ivan Vogelius, Ph.D.
- Associate professor Marianne Aznar, Ph.D.
- Associate professor Per Nilsson, Ph.D.

Submitted: April 2nd 2013
Thesis title: Risk-based optimization of photon and proton radiotherapy for pediatric medulloblastoma

Author: Patrik Brodin

Name of departments: Department of Radiation Oncology
Rigshospitalet
Blegdamsvej 9
DK - 2100 Copenhagen

Niels Bohr Institute
University of Copenhagen
Universitetsparken 5
DK - 2100 Copenhagen

Principal supervisor: Associate professor Per Munck af Rosenschöld, Ph.D.

Academic advisors: Associate professor Thomas Björk-Eriksson, M.D., Ph.D.
Medical physicist Ivan Vogelius, Ph.D.
Associate professor Marianne Aznar, Ph.D.
Associate professor Per Nilsson, Ph.D.

The work in this thesis was funded by grants from the Danish Child Cancer Foundation

Submitted April 2nd 2013 to the PhD School of Science, Faculty of Science, Niels Bohr Institute, University of Copenhagen for the PhD degree
To my mother and father, for raising me to have a curious mind always seeking to learn new things.

"Study hard what interests you the most in the most undisciplined, irreverent and original manner possible."

- Richard P. Feynman
CONTENTS

ABSTRACT ........................................................................................................................................... VII

DANSK RESUMÉ (DANISH SUMMARY) ................................................................................................ VIII

LIST OF PAPERS .................................................................................................................................. X

ACKNOWLEDGEMENTS ....................................................................................................................... XI

ABBREVIATIONS .................................................................................................................................. XIII

BACKGROUND ...................................................................................................................................... 1

THE DEVELOPMENT OF RADIOTHERAPY FOR MEDULLOBLASTOMA .............................................. 3
RADIOTHERAPY DOSE AND USE OF CHEMOTHERAPY ..................................................................... 6

AIMS OF THE THESIS ............................................................................................................................. 8

RESEARCH EFFORTS TO LIMIT THE RISK OF SEVERE LATE COMPLICATIONS ......................... 9

Proton radiotherapy for medulloblastoma ............................................................................................ 9
The problem of comparing different long-term complication risks .................................................... 10
Reducing the risk of neurocognitive decline ....................................................................................... 12
Reducing the high-dose treatment volume ......................................................................................... 14

RESEARCH EFFORTS TO IMPROVE TUMOR CONTROL .................................................................... 17

Altering the radiotherapy fractionation ............................................................................................... 17
Developing a mathematical tumor control model for medulloblastoma ............................................. 19

SIMULTANEOUS OPTIMIZATION OF TUMOR CONTROL AND TOXICITY ...................................... 22

SPECIAL APPLICATIONS OF STATISTICAL METHODS IN THIS THESIS ....................................... 26

SUMMARY AND CONCLUSIONS .......................................................................................................... 28

FUTURE PERSPECTIVES IN MEDULLOBLASTOMA RESEARCH ...................................................... 29

REFERENCE LIST .................................................................................................................................. 32

PAPERS I - VI
ABSTRACT

Medulloblastoma (MB) is one of the most common brain tumors in children and most patients will survive their disease if treated with surgery, radio- and chemotherapy. This has resulted in a growing number of MB survivors who are at risk of developing severe late complications related to the aggressive anti-cancer therapy. Some of the most important late complications for these patients include secondary malignancies, cardiac events and neurocognitive decline. Much of the late complication risk can be attributed the craniospinal irradiation which is part of standard management, since this results in irradiating a large volume of healthy tissue in these young patients. The general aim of this thesis is reducing the risk of late treatment toxicity through risk-based advanced radiotherapy optimization. We specifically investigated the use of radiobiological models to estimate late complication risks, developed a mathematical framework for comparing different risks on a common scale and the assessed the potential clinical benefit of sparing critical neurological structures during cranial irradiation. We also developed a mathematical tumor control model for standard-risk MB to assess the effect on treatment efficacy from applying different risk-based treatment strategies.
DANSK RESUMÉ (DANISH SUMMARY)


Ligeledes udviklede vi en metode til at sammenligne risikoen for de enkelte senkomplikationer ved at anslå ”tabte leveår” som følge af de forskellige senkomplikationer. Dette tillader en direkte sammenligning mellem f.eks. risikoen for hjertesvigt og risikoen for lungekræft som følge af strålebehandling, hvilket ellers er meget svære at sammenligne. Dog kan alle senkomplikationer ikke måles i tabte leveår, som f.eks. nedsatte kognitive evner, der uheldigvis er meget hyppig efter helhjerne-bestråling. Vi fandt, ved at udlede modeller for sammenhæng mellem dosis og effekt, at ved at reducere stråledosis til kritiske områder i hjernen, såsom hippocampus, ville den anslåede risiko for senere tab af kognitive evner blive formindsket betragteligt. Også til dette vurderedes proton-bestråling en langt bedre teknik.

LIST OF PAPERS

Study I

Study II

Study III
*Equal contribution

Study IV

Study V

Study VI
Brodin NP, Vogelius IR, Björk-Eriksson T, Munck af Rosenschöld P, Maraldo MV, Aznar MC, Specht L, Bentzen SM. Risk-based radiation therapy optimization through simultaneous common scale comparison of the life years lost attributable to tumor control and late complication risk. (Manuscript)
ACKNOWLEDGEMENTS

Firstly, my most sincere thanks go to my supervisors for supporting me, helping me and guiding me along the way, without you this work would not have been completed. I wish to thank Per Munck af Rosenschöld, my principle supervisor, for his leadership, his many ideas and his unprecedented ability to make me push the limits of my own abilities. He is responsible for a great deal of the innovative thinking and high quality of the studies included in this thesis. Thomas Björk-Eriksson for being not only a fountain of clinical knowledge but easily one of the best supervisors a PhD student can wish for. Thomas’ advice, professional and personal, has kept my spirits high and my academic curiosity at its peak throughout the entirety of my PhD studies. Ivan Vogelius for a countless number extremely fruitful statistical and mathematical discussions, making sure all of the analyses conducted in this thesis were of the highest quality. Also, on a personal note, I owe Ivan thanks for three great months working at the University of Wisconsin, Madison, USA, one of the best experiences of my career so far. I also owe a great deal of thanks to Marianne Aznar who taught me almost all I know about treatment planning and how to be a good researcher. Marianne has been there for me since day one and I will always appreciate that. I wish to thank Per Nilsson for his enormous cunning within the field of medical physics, many of the innovative ideas for the work in this thesis comes directly from him.

Thanks to professor Lena Specht for showing such a great interest in my work and sharing her vast experience in the field of radiotherapy research with me. Professor Svend-Aage Engelholm for his confidence in me and the work I do and for our many interesting discussions. Professor Søren Bentzen for his knowledge, great ideas and impressive grasp of medical statistics as well as for making my exchange visit in Madison such a great experience. Anne-Kiil Berthelsen for spending an enormous amount of time meticulously delineating tumors and risk organs, without her none of the work in this thesis would have been possible. Lars Ohlhues and Søren Lassen for helping me with technical tasks such as extracting data from the radiotherapy database. Thomas Carlslund and Mikael Olsen for providing a ton of vital technical support. Karin Andreasen and Tove Romer for helping me with all kinds of administrative issues and answering all of my questions with a smile.
Thanks to Dr. Maja Maraldo for all of her input on clinical questions, scientific writing, helping me write the Danish summary of the thesis and for all the great work we have done together. Thanks also for being a great friend and making sure, together with Dr. Søren Møller, that the time spent in our little office was the best three years so far. Thanks to Tobias Pommer and Daniel Gasic, two of my best friends whom I have the pleasure of also calling my colleagues, for endless trips on the train to and from work and discussions about science, programming, physical exercise, food, wine, and everything else.

Thanks Christian, Mikael, Joen, Henrik, Malene, Jacob, Josefine, Troels, Anders, Katrin, Jonas, Gitte, Ditte, Anne, Marianne and any other past or present members of the interdisciplinary research group at the Department of Radiation Oncology. Thanks also to all physicists, physicians and radiographers working in the clinic for sharing your experience and knowledge with me. I would also like to thank all of my friends, who have kept me sane and socially functioning throughout my PhD studies.

Lastly, I would like to thank Tony Christie’s “Is this the way to Amarillo” and the AVGN YouTube videos for keeping me company during many long nights writing this thesis.

The work presented in this thesis was made possible through grants from the Danish Child Cancer Foundation.
ABBREVIATIONS

CNS  Central nervous system
PNET  Primitive neuroectodermal tumor
MB  Medulloblastoma
CSI  Craniospinal irradiation
Gy  Gray
PF  Posterior fossa
CSF  Cerebrospinal fluid
PFS  Progression-free survival
kV  Kilovoltage
2D  Two dimensional
CT  Computed tomography
3D CRT  Three-dimensional conformal radiotherapy
MRI  Magnetic resonance imaging
IMRT  Intensity-modulated radiotherapy
VMAT  Volumetric-modulated arc therapy
OED  Organ-equivalent dose
CCSS  Childhood Cancer Survivor Study
LYL  Life years lost
QALYs  Quality-adjusted life years
DALYs  disease-adjusted life years
IQ  Intelligence quotient
SVZ  Subventricular zone
OR  Odds ratio
PTV  Planning target volume
GTV  Gross tumor volume
CTV  Clinical target volume
BED  Biological effective dose
LQ  Linear-quadratic
EQD2  Equivalent dose in 2 Gy fractions
EFS  Event-free survival
SIOP  International Society of Paediatric Oncology
TCP  Tumor control probability
FFP  Freedom from progression
Th1-Th10  The 1st to 10th thoracic vertebra
CI  Confidence interval
MC  Monte Carlo
**Background**

Second only to leukemia, tumors of the central nervous system (CNS) are the most commonly occurring cancers in children.\(^1\) Approximately 20% of childhood CNS malignancies are primitive neuroectodermal tumors (PNETs), most of which present as medulloblastoma (MB) which presents in the posterior cranial fossa.\(^2\) There are about 10 new MB cases per year in Denmark and correspondingly 400 new cases per year in the US.\(^2-4\) Research into MB has been going on for the better part of 80 years starting with the pioneering work of neurosurgeon Harvey Cushing who diagnosed, operated on and carefully monitored these patients.\(^5\) The chance of survival for MB remained abysmal when applying surgery as the only treatment. The most considerable breakthrough in MB treatment came with the introduction of craniospinal irradiation (CSI) in the 1950s. This treatment strategy was proposed as a result of metastases in the brain and spinal cord found on post-mortem examination.\(^6,\)\(^7\) By treating the whole craniospinal axis to a radiation dose of about 35 Gray (Gy), delivered in fractions of 1.6 Gy per day, and further boosting the posterior fossa (PF) to a dose of 50 Gy, almost two thirds of patients went on to become at least 3-year survivors.\(^6\) Compelling evidence of the importance of CSI to complement the focal PF treatment was presented from the Lund University hospital in a review of the MB treatment at their institution from 1946 to 1975.\(^8\) They presented an increase in 10-year survival from approximately 5% to 25% when not only treating the PF but also including irradiation of the spinal axis. The second considerable improvement in the treatment of this disease was the introduction of adjuvant chemotherapy in the late 1970s and 1980s.\(^9,\)\(^10\) The wide acceptance of chemotherapy as a beneficial addition to surgery and

![Figure 1. A primary medulloblastoma (red contour) and posterior fossa (pink contour).](image)
radiotherapy was, however, far from immediate, with some promising but non-conclusive results from the initial randomized trials.\textsuperscript{11-14}

One of the most important contributions from the large number of randomized MB trials is the many prognostic factors that have been identified to strongly correlate with treatment outcome and survival. Patient characteristics found to considerably impact the prognosis of MB patients are age at diagnosis, gender, presentation of radiologically detectable metastases within the CNS, involvement of the cerebrospinal fluid (CSF), disease stage and the extent of surgical resection.\textsuperscript{11, 13-17} This has led to the widespread consensus of classifying patients with MB according to different risk groups, based on updates of the classical Chang staging system for MB.\textsuperscript{18, 19} Patients with no detectable disease outside of the PF, either via radiological examination or examination of the CSF, older than 3 years and without any post-operative residual tumor larger than 1.5 cm\textsuperscript{2} are considered standard-risk. Patients presenting with radiological metastases, CSF involvement or large residual tumors are classified as having high-risk disease and a worse prognosis.\textsuperscript{20-24} Very young MB patients (< 3 years old) are considered separately, unfortunately also with a poor prognosis due to a likely combination of more aggressive disease and lower treatment tolerance compared to older children.\textsuperscript{14, 20, 25} In this very young patient group the focus has been to defer from CSI for as long as possible through the use of chemotherapy with or without local PF irradiation.\textsuperscript{20, 26, 27}

This thesis focuses on standard-risk MB for whom the prognosis with current multimodality treatment is quite promising with progression-free survival (PFS) rates of 75-85\%.\textsuperscript{28, 29} Achieving such high survival rates has resulted in a growing population of long-term pediatric MB survivors, at risk of suffering adverse late effects induced by the aggressive anti-cancer treatment. These young patients are at risk of developing growth retardation, loss of hearing or vision, neurocognitive decline, endocrine deficits, cardiac- or pulmonary complications and secondary cancers as late occurring complications.\textsuperscript{30-49}
The development of radiotherapy for medulloblastoma

Since the 1950s when CSI was introduced for pediatric MB patients, the technical development of radiotherapy has been closely linked with its application in MB treatment. Originally, MB patients were treated with orthovoltage (250 kV) photons (x-rays) guided only by patient thickness and certain anatomical landmarks, and the absorbed radiation dose was calculated manually using depth dose tables.\textsuperscript{6, 21, 50} The introduction of 2D radiographs into treatment planning enabled simulating the set-up of radiotherapy fields on the individual patient. A further treatment advance was the use of high-energy \textsuperscript{60}Cobalt gamma rays instead of orthovoltage photons, which could treat the target to the desired dose without the very high skin doses related to the earlier technique.\textsuperscript{51} Currently, all external beam radiotherapy is delivered using linear accelerator technology where megavoltage photons are aimed at the target volume through a gantry delivery system.

The classical field set-up for 2D treatment planning in MB was two opposed lateral fields covering the whole brain and part of the cervical cord, matched with one or two spinal posterior fields depending on patient height. The PF boost was classically treated with two opposed lateral fields.\textsuperscript{51-53} A key feature for accurate CSI is the matching between adjacent fields where the field junctions are typically "feathered" (moved 1 cm every 6 or 7 fractions) to decrease the risk of considerable under- or overdosage locally due to the difficulty in exact field matching. Also sufficiently treating the cribriform plate (located in the frontal lobe between the eyes) while sparing the eyes from high radiation doses is another complicating feature of CSI.\textsuperscript{52, 53}

A major technical breakthrough in the radiotherapy of MB (and radiotherapy in general) was the introduction of computed tomography (CT) simulation providing 3D anatomical data for treatment field set-up.\textsuperscript{54} More importantly, the use of CT scans enabled computer-driven 3D dose calculations and the possibility to apply 3D conformal radiotherapy (3D CRT) for which the treatment fields could be set to accurately cover the target volume while shielding key sensitive structures.\textsuperscript{55, 56} Here, a multi-leaf collimator positioned on the head of the treatment gantry is used to shape the radiation field to fit the target volume.
Figure 2 illustrates a 3D CRT set-up for a MB patient with two opposed cranial fields, one posterior spinal field and a four field (two opposed and two oblique) PF boost.

**Figure 2.** Field setup for craniospinal irradiation and high-dose boost using 3D conformal radiotherapy.

In the last few decades, the CT treatment planning for MB has been complemented by magnetic resonance imaging (MRI) often performed using a Gadolinium contrast agent. The superior soft tissue contrast and high anatomical resolution renders MRI the preferred tool for determining accurate diagnosis and highlighting critical intra-cranial structures such as the cochlea, optic chiasm, brainstem, hypothalamus and pituitary gland.

Intensity-modulated radiotherapy (IMRT), another technical radiotherapy advance has, however, not been widely implemented for MB patients. IMRT provides a computer-driven inverse-planned optimization of radiation treatment plans. The IMRT principle is based on inversely optimizing the pattern of dose delivery from a number of incident treatment fields to yield a predetermined dose distribution within the patient. This allows for tailoring the radiation dose distribution around critical structures while successfully treating complex target volumes. Since IMRT often results in a large volume of healthy tissue being irradiated to low doses, the use in pediatric patients has been limited due to the concern of increasing the risk of developing radiation-induced secondary cancers. Even more recent advances in radiotherapy have led to the implementation of rotational IMRT through either volumetric-modulated arc therapy (VMAT) or helical
As for static-field IMRT these techniques also utilize inverse-planned radiation delivery although the dose is delivered while the treatment machine is continuously rotating around the patient.

![Dose distributions in the sagittal and transversal plane from craniospinal irradiation, shown as dose color-wash.](image)

**Figure 3.** Dose distributions in the sagittal and transversal plane from craniospinal irradiation, shown as dose color-wash. Reprinted from Brodin et al. with permission from Acta Oncologica.

Treating pediatric patients with proton radiotherapy is expected to have a major impact on reducing the dose to healthy tissues, especially for patients undergoing CSI. Protons deposit their energy in a different way compared to high-energy photons, allowing for very sharp dose gradients within the patient. This is especially true at the distal edge of the proton path where most of the energy is deposited within a very short range, referred to as the Bragg peak. Proton radiotherapy can be delivered according to two different principles, either using passive scattering or active spot-scanning. In passive scattering delivery a collimator block designed to the specific patient case is positioned in the beam path to shape the resulting absorbed dose within the patient to the target volume. Spot-scanning utilizes magnets to actively steer the proton beam during delivery and thus tailor the absorbed dose to the target volume.
Radiotherapy dose and use of chemotherapy

The search for the appropriate radiation dose for MB patients has been the subject of intense research for the past 30 years. Regarding the primary tumor boost, however, it has been well established that a dose above 50 Gy is necessary to provide an effective treatment, as summarized in a review by Paulino in 2002 (cf. Table 5 in this reference). For the craniospinal dose it is a different story, where the standard prescribed dose was originally 36 Gy for all patients. However, the severe neurological and neuroendocrine complications observed after this treatment warranted efforts to reduce the CSI dose. This resulted in risk-stratified treatment approaches attempting to reduce the CSI dose in standard-risk patients to 23-24 Gy while treating high-risk patients to a higher dose of ~38-40 Gy with a boost to any neuraxial metastases up to 45 or 50 Gy. The search for the appropriate CSI dose in the different risk groups led to a number of randomized trials also investigating the role of chemotherapy in a risk-stratified setting.

High survival rates were maintained in standard-risk patients when lowering the CSI dose to 23.4 Gy if supplemented by post-irradiation chemotherapy consisting of vincristine, CCNU (or replacing CCNU with cyclophosphamide) and cisplatin. Other chemotherapy agents have also been tested such as the “eight-drugs-in-one-day” regimen, although this compared unfavorably to a vincristine and CCNU based regimen. Chemotherapy does play an important role in high-risk disease although here in combination with high-dose radiotherapy. The timing of radiotherapy and chemotherapy also has a large impact on the success of MB treatment. It has been shown that radiotherapy should be started as soon as possible following surgery and the duration of radiation treatment should be kept < 50 days. Consequently, post-irradiation (maintenance) chemotherapy is more efficient compared to adjuvant pre-radiation chemotherapy since this is likely to prolong the start of radiotherapy and overall treatment time.
For standard-risk pediatric MB the current standard of care in most centers is surgery followed by craniospinal photon radiotherapy of 23.4 Gy and a PF boost up to 54-55.8 Gy, delivered in 1.8 Gy fractions. Chemotherapy is administered as concomitant vincristine during radiotherapy followed by post-irradiation maintenance chemotherapy with vincristine, CCNU and cisplatin. The risk-based optimization strategies explored in this thesis should therefore be compared against this standard.
Aims of the thesis

The overall aim of this PhD thesis was to develop mathematical models and decision making tools in order to make informed decisions based on risk estimates during planning of radiotherapy, and to quantify the potential clinical benefit of risk-based optimization of modern radiotherapy techniques for standard-risk MB patients.

Specific aims:

1. To review the published literature in order to find dose-response data for late radiotherapy complications in pediatric cancer patients. Also to use the available data to construct mathematical dose-response models for adverse late effects such as secondary cancers, cardiac complications and neurocognitive decline for pediatric MB patients. (Study I, Study III)

2. To develop a mathematical framework for comparing different types of late complication risks on a common scale, accounting for differences in the time-to-event distribution and effects of varying age at exposure. (Study II)

3. Developing strategies to spare the neurogenic niches in pediatric MB patients from high doses of radiation comparing different radiation modalities. In addition, to evaluate the estimated clinical benefit of sparing the neurogenic niches by way of mathematical neurocognitive dose-effect models. (Study III, Study IV)

4. To develop a tumor control dose-response model for standard-risk MB patients. Also, to provide a framework, using this tumor control model and the previously developed models for long-term risk estimation, for directly quantifying the trade-off between tumor control and treatment-related toxicity. (Study V, Study VI)
Research efforts to limit the risk of severe late complications

Proton radiotherapy for medulloblastoma

The prospect of craniospinal proton therapy is very attractive as the sharp dose gradient at the distal edge of the beam would allow for sparing most of the ventral part of the thorax. This has been elegantly shown on MRI scans of children treated with craniospinal proton irradiation. The edge of the proton beam across the vertebral bodies in the treatment planning system coincided very well with changes in the T1-weighted MR signal as a result of radiation-induced fat infiltration of the bone. This offers a great dosimetric advantage over photon treatment which has been shown in several treatment planning studies.

Although there is little clinical evidence to date for pediatric patients treated with proton therapy one exception is a recent report on MB patients showing low rates of high-grade hearing loss 1 year after proton radiotherapy.

Secondary cancers is an increasing concern for pediatric patients treated with radiotherapy and the potential for using proton therapy to reduce the secondary cancer risk has received much attention. To this end, we attempted to quantify the reduction in secondary cancer risk, as well as the potential risk reduction for other non-cancer late complications, for pediatric MB patients treated with either 3D CRT, VMAT or proton therapy (Study I). Here, we estimated the secondary cancer risk based on an organ-equivalent dose (OED) model with a plateau dose-response relationship, i.e. the risk increased with dose until a certain level where it converges. This concept was introduced by Schneider et al. and the details of our implementation are given in (Study I).

We estimated the risk of inducing a solid secondary cancer at an attained age corresponding to the average lifetime of a Danish person (78.5 years) to be 45%, 56% and 7% for 3D CRT, VMAT and proton therapy, respectively for a prescribed CSI dose of 23.4 Gy. This showed a clear advantage of protons compared to photons although our estimates were somewhat high compared to previous treatment planning studies which estimated risks of about 20-55% for MB patients treated with IMRT. When estimating the risk of developing a secondary cancer up to a certain attained age, this has to be
conditional on the patient surviving until that age. Our risk estimates were conditioned on the survival probability of a person in the general population which is likely a source for overestimation since the long-term survival of a pediatric cancer patient is considerably less.

Also, the potential carcinogenic effect of secondary neutrons attributable to proton beam scattering has to be considered. The secondary neutron contribution is estimated to be much less using a spot-scanning proton delivery system compared to passive scattering.\textsuperscript{79, 86} A group at the MD Anderson Cancer Center has used Monte Carlo methods for simulating the neutron contribution on mathematical phantoms resembling pediatric patients, which can then be included in the risk calculations.\textsuperscript{79, 83, 86} Considering the estimated amounts of secondary neutrons produced by proton beams, especially in the spot-scanning setting, it is unlikely that this would render proton therapy dosimetrically unfavorable compared to photons for craniospinal treatment.

\textit{The problem of comparing different long-term complication risks}

The Childhood Cancer Survivor Study (CCSS) is a large collaborative effort in the US that have followed approximately 15,000 pediatric cancer patients, surviving at least 5 years after diagnosis, for almost 30 years now. This has been an important source of knowledge when it comes to late complications for pediatric cancer patients, both in terms of secondary cancers and non-malignant adverse events.\textsuperscript{38, 47, 87-89} It is clear that survivors of pediatric MB are at risk of many kinds of late complications including cardio-pulmonary effects, secondary cancers, endocrine deficits and neurocognitive decline. Since these patients are at risk of many different late events it can be difficult to estimate the extent to which these complications affect MB survivors. A common scale measure comparing different types of late complications could provide a helpful decision making tool when comparing e.g. different treatment options.

By estimating the life years lost (LYL) attributable to different late complications these can be compared directly in a measure which is easy to interpret. In Study II we estimated the
LYL from secondary lung, breast, thyroid and stomach cancer as well as myocardial infarction and heart failure, for a group of MB patients. This enabled us to directly compare the LYL between 3D CRT, VMAT and proton therapy, visualizing the difference in estimated complication risk in a single measure. Here we found a clear advantage for proton therapy because of the low dose to the lungs, heart and breasts. For 3D CRT compared to VMAT there was a clear trade-off between the increased risk of secondary cancer through the large volume receiving a low dose with VMAT compared to a higher dose to the heart with 3D CRT, leading to an increased risk of cardiac complications.

The LYL measure has the advantage that the effects of age at exposure and time-to-event are included in the estimation. This weighs early occurring events higher as there are more life years at stake, which is also the case for younger age at exposure. To account for competing risks of death when estimating the LYL attributable to each complication, we used the conditional survival probability, derived from long-term survival data of pediatric cancer patients\textsuperscript{90}, in contrast to using the general population survival as in Study I. The LYL measure can be used to compare different treatment modalities but also to estimate whether re-optimizing treatment plans within the same modality leads to a reduced or increased late complication burden.

A natural although challenging extension to the LYL measure would be to include also non-lethal complications such as hypothyroidism, hearing loss or cognitive decline as these have considerable impact on the quality of life for a cancer survivor. Estimating the quality-adjusted life years (QALYs) lost would in that sense provide a better position for clinical decision making as it considers all treatment complications. How to weight the non-lethal complications to yield the respective QALYs is, however, difficult and may vary considerably between patients and type of primary cancer. Attempts have been presented e.g. by Smith et al. who presented a Markov model for estimating the quality-adjusted life expectancy after prostate cancer treatment.\textsuperscript{91} An interesting mathematical framework has also been developed to estimate the disease-adjusted life years (DALYs) lost attributable to different cancers on a nationwide scale.\textsuperscript{92} Combined with models for estimating normal tissue complications, this may prove a useful approach also for individualized treatment evaluation.
Reducing the risk of neurocognitive decline

Neurocognitive decline is unfortunately both a common and devastating complication after whole-brain irradiation of MB patients. This is one of the main reasons for lowering the CSI dose to 23.4 Gy in standard-risk patients although many patients still present with reduced cognitive function after MB treatment. Merchant et al. developed dose-response models for evaluating the IQ decline in a prospective evaluation of 39 pediatric CNS patients undergoing CSI.\textsuperscript{9} They investigated dose-volume effects in different parts of the brain and found that IQ decline correlated with large volumes receiving high doses, almost independent of which part of the whole brain was considered. In a subsequent analysis, Merchant et al. evaluated whether proton therapy was estimated to lower the risk of cognitive decline compared to photon treatment for a number of pediatric CNS patients.\textsuperscript{93} Here they found that proton therapy reduced the volumes of the supratentorial brain and temporal lobes receiving low and intermediate doses, which translated into less estimated IQ decline. For MB patients, however, the whole-brain irradiation component dampened the differences between proton and photon treatment in their analysis, suggesting that proton therapy would not considerably reduce the risk of cognitive decline in this patient group.

Recently the neurogenic niches have attracted much attention as potential critical structures for radiation-induced neurocognitive decline. In mammals, the neurogenic niches are located in the denate gyrus of the hippocampus and the subventricular zone (SVZ) of the lateral ventricles.\textsuperscript{94, 95} For MB patients it may be difficult to reduce the risk of cognitive impairment by lowering the dose to larger brain components since whole-brain irradiation is part of standard management. Thus, we analyzed the potential of neurocognitive sparing radiotherapy for cranial irradiation in MB by limiting the dose to the neurogenic niches (Study III). To estimate the clinical benefit of sparing the neurogenic niches we used neuropsychological outcome data from long-term survivors of pediatric CNS malignancies.\textsuperscript{30} Through retrospective evaluation of radiation treatment records, the authors found a strong correlation between cognitive impairment and the dose to the temporal lobe. Using these data, with the assumption that sparing the
neurogenic niches equals sparing the whole temporal lobe, we derived dose-response models for the risk of various cognitive impairments according to a logistic relationship:

\[
OR_D = \frac{p_D / (1-p_D)}{p_0 / (1-p_0)}
\]

\[
OR_D = OR_{10}^{\frac{n_{00}}{n_{DD}}}
\]

\[
\Rightarrow p_D = \frac{OR_{10}^{\frac{n_{00}}{n_{DD}}}}{1 + \frac{OR_{10}^{\frac{n_{00}}{n_{DD}}}}{p_0}}
\]

(Eq. 1)

where \(D\) is the dose in Gy, \(OR_{10}\) is the corresponding odds ratio at 10 Gy, \(p_0\) is the baseline risk of impairment at zero dose and \(p_D\) is the risk of impairment at dose \(D\). The steepness of the dose-response curve being determined by the \(OR_{10}\) in Armstrong et al.\(^{30}\) for impaired task efficiency, memory or organization. Details of the dose-response models and applied assumptions are given in Study III although it is interesting to compare this model to Merchant’s model for IQ decline as a function of temporal lobe dose\(^{39}\):

\[
IQ = \text{Intercept} + \left( \beta_{\text{age}} \cdot e - \beta_{V_{15-40Gy}} \cdot V_{15-40Gy} - \beta_{V_{40-65Gy}} \cdot V_{40-65Gy} \right) \cdot t
\]

(Eq. 2)

where the \(\beta\) values represent the corresponding dose-response parameters and in contrast to our model this also considers the age at exposure \((\varepsilon)\) and the time after treatment, \(t\), whereas our models estimate only long-term (up to ~25 years) risks of impairment. Also our model has a continuous dose-response which is based on mean dose, whereas their model for IQ decline is based on how much of the temporal lobe volume is irradiated to either an intermediate dose of 15-40 Gy or a high dose of 40-65 Gy.

Using our dose-response models and treatment plans optimized with the aim of limiting the dose to the hippocampus and the SVZ, we showed that a considerable reduction in estimated risk of cognitive impairment was possible when using intensity-modulated proton therapy. Some risk reduction was possible using photon IMRT or VMAT but considerably less compared to proton therapy. In these treatment plans we deliberately underdose structures which have previously been part of the whole-brain target volume.
Although the volume of the hippocampus and the SVZ make up only about ~1% of the whole brain volume, the potential increase in recurrence risk from this underdosage must be weighed against the reduced risk of cognitive impairment.

*Reducing the high-dose treatment volume*

Another effort that appears promising for limiting neurocognitive toxicity in MB patients is to reduce the volume of the high-dose boost target, typically treated to 54-55.8 Gy. This is achieved by moving away from whole PF irradiation and only treating the pre-operative tumor bed including a margin. Although there appears to be a lack of consensus for how large this margin should be or how to exactly define the tumor bed on MRI scans, several prospective cohort studies have shown good posterior fossa control rates using this treatment approach, cf. Table 1. Not all of these studies included cognitive evaluation of the patients, but two studies reported cognitive functioning that compared favorably to standard treatment. This effect may, however, be attributed the fact that these studies investigated treatment without chemotherapy. Reducing the volume of the boost target provides the potential of further limiting the dose to critical neurological structures compared to whole PF treatment. The choice of treatment margin also has a considerable effect on the resulting dose to these structures, mainly the hippocampus, as the margin controls the size of the final planning target volume (PTV).

The role of the hippocampus as a key neurological risk organ in pediatric cranial irradiation is becoming increasingly clear although the importance of the SVZ in cognitive functioning remains under debate. This was most recently illustrated in a report by Redmond et al. where they prospectively monitored the cognitive function in 19 pediatric patients with CNS malignancies. They found significant correlations between the dose to the hippocampus and temporal lobe and decreased motor speed and dexterity, whereas no significant correlation was found between SVZ dose and cognitive function.
Table 1. Prospective studies of conformal tumor bed boost for medulloblastoma. The gross tumor volume (GTV) is given by the residual tumor and the resection cavity. The clinical target volume (CTV) and planning target volume (PTV) are expanded from the GTV.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Boost dose</th>
<th>GTV</th>
<th>CTV</th>
<th>PTV</th>
<th># patients</th>
<th>5-year PF control rate</th>
<th>5-year tumor bed control rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrie et al. (JCO 2009)</td>
<td>68 Gy (no chemo)</td>
<td>Tumor bed by MRI + 1.5 cm</td>
<td>N/A</td>
<td>N/A</td>
<td>48</td>
<td>N/A</td>
<td>89.6%</td>
</tr>
<tr>
<td>Douglas et al. (IJROBP 2004)</td>
<td>55.8 Gy</td>
<td>Tumor bed by MRI</td>
<td>GTV + 1.0 cm</td>
<td>CTV + 0.5 cm</td>
<td>33</td>
<td>94%</td>
<td>94%</td>
</tr>
<tr>
<td>Gupta et al. (IJROBP 2012)</td>
<td>68 Gy (no chemo)</td>
<td>Tumor bed by MRI</td>
<td>GTV + 1-1.5 cm</td>
<td>CTV + 0.5 cm</td>
<td>20</td>
<td>95%</td>
<td>95%</td>
</tr>
<tr>
<td>Polkinghorn et al. (IJROBP 2011)</td>
<td>54-55.8 Gy</td>
<td>Tumor bed by MRI</td>
<td>GTV + 1.0 cm</td>
<td>CTV + 0.5 cm</td>
<td>33</td>
<td>84.8%†</td>
<td>84.8%†</td>
</tr>
<tr>
<td>Wolden et al. (JCO 2003)</td>
<td>55.8 Gy</td>
<td>Tumor bed by MRI</td>
<td>GTV + 1-2.0 cm†</td>
<td>CTV + 0.5 cm</td>
<td>32</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

†NB: some high-risk patients in this cohort
††Margin selected so CTV did not extend outside the bony confines of the PF

We investigated the potential clinical benefit of planning the high dose boost in MB patients using a hippocampal sparing treatment approach with either IMRT or proton therapy (Study IV). We also evaluated the effect of applying different treatment margins, resulting in larger or smaller target volumes. Although the whole-brain radiotherapy target is very similar between patients, there is considerably more inter-patient heterogeneity in the size and position of the boost target. To this end, the primary tumors and hippocampi where delineated on transversal T1-weighted MRI scans of 17 pediatric MB patients. The relative size and position of these structures for this group of patients are illustrated in Figure 4, in relation to the PF. Detailed descriptions of the treatment planning and risk calculations are provided in Study IV.
In this analysis we found that there can be a considerable dose contribution to the hippocampus from the boost treatment which depended both on the size of the primary tumor and on the margin applied. The tumor bed was located along the central axis of the posterior fossa for most of the patients. The largest reduction in hippocampal dose and corresponding risk of cognitive impairment was attributed the use of hippocampal sparing radiotherapy. Both IMRT and proton therapy could significantly reduce the hippocampus dose, although protons performed better in this setting. Interestingly, we also found that the distance between the closest point of the PTV and the center of the hippocampus appears to be an appropriate surrogate for predicting the hippocampal dose, at least for distances up to 2 cm. This allows for extraction of a simple distance measure to evaluate whether a hippocampal sparing approach would be beneficial and how much the dose could be lowered with IMRT or proton therapy compared to conventional 3D CRT. This could aid in selection of patients for hippocampal sparing radiotherapy without the lengthy process of treatment plan comparison. Hippocampal sparing may be a promising method for reducing the risk of late neurocognitive decline in MB patients, especially when applied both in the CSI part and the boost treatment, cf. results section in Study III, IV.
Research efforts to improve tumor control

Altering the radiotherapy fractionation

As mentioned in the introduction the focus of this thesis is risk-based comparison of treatment regimens aiming to reduce the risk of late complications in standard-risk MB survivors. There have also been recent efforts to improve the tumor control rate in these patients. The therapeutic ratio, i.e. chance of controlling the tumor without inducing severe complications, can be increased by altering the radiotherapy fractionation, based on the different recovery times for the cancer cells and surrounding healthy tissues. The maximal therapeutic ratio is reached when the dose per fraction is set so that the tumor recovers as little as possible until the next fraction, while not exceeding the tolerance dose for the irradiated healthy tissue. A well-established mathematical relation for describing the different biological effective dose (BED) between different fractionation schemes is the linear-quadratic (LQ) model, which is originally based on surviving fractions of irradiated cell cultures. The model gives the surviving fraction as:

$$S = e^{-(\alpha D + \beta D^2)}$$ \hspace{1cm} (Eq. 3)

where $\alpha$ and $\beta$ are tissue dependent constants and $D$ is the radiation dose. Through this relation the BED can be written as:

$$BED = nd \left(1 + \frac{d}{\alpha / \beta}\right)$$ \hspace{1cm} (Eq. 4)

where, $n$ is the number of treatment fractions, $d$ is the dose per fraction and $\alpha / \beta$ the parameter that controls the fractionation sensitivity of the given tissue. It is the difference in $\alpha / \beta$ between tumor and healthy tissue that drives optimizing the fractionation to give the highest BED for the tumor while keeping the BED for healthy tissues as low as
possible. Unfortunately, most $\alpha/\beta$ values are subject to large uncertainty when it comes to the clinical in vivo setting.

For pediatric MB, it has been suggested that reducing the fraction size from the standard 1.8 Gy and treating patients twice daily in a so called hyperfractionation regimen could increase tumor control rates without increasing rates of normal tissue toxicity. Two major investigations for hyperfractionated treatment of standard-risk MB have been undertaken. One is a French study where patients were treated with hyperfractionated radiotherapy and no chemotherapy, which aimed to increase the therapeutic ratio by eliminating the chemotherapy-related toxicity. Patients were treated with CSI of 36 Gy in 1 Gy per fraction and a tumor bed boost up to 68 Gy, also in 1 Gy per fraction. Using the LQ-model, this corresponds to an equivalent dose in 2 Gy fractions (EQD2) of 62 Gy to the tumor bed and 33 Gy to the craniospinal axis when considering the effect on the tumor (assuming $\alpha/\beta = 10$). For the standard treatment the corresponding EQD2 is 53 Gy and 23 Gy, respectively. The EQD2 for the healthy CNS tissue was estimated to be 51 Gy and 27 Gy in the hyperfractionated setting compared to 52 Gy and 22 Gy in the standard regimen, estimated with $\alpha/\beta = 2$ for the healthy tissue. The aim of this higher BED compared to the standard treatment was to maintain a high rate of tumor control when avoiding the addition of chemotherapy. This study showed promising long-term results with a 6-year event-free survival (EFS) of 75% and IQ decline that compared favorably to previous reports for the standard treatment regimen.

Another important investigation of hyperfractionation for MB was the SIOP PNET-4 randomized trial which compared the standard treatment regimen to the same hyperfractionation regimen as in the French study. Here, chemotherapy was included in both treatment arms and the study hypothesis was that the tumor control would be higher in hyperfractionated arm with a higher biologically effective tumor dose. With 179 patients enrolled in each treatment arm, they found no difference in 5-year EFS between the two regimens. This raises the question of whether the dose-response relation for MB tumor control is different than what was expected. Since the introduction of chemotherapy, this
is the first MB trial directly comparing two different radiotherapy regimens, suggesting that the dose-response relation established in the pre-chemotherapy era may not be valid when adjuvant chemotherapy is included in the treatment.

**Developing a mathematical tumor control model for medulloblastoma**

Because of this apparent lack of knowledge for the tumor control when chemotherapy is included in the treatment, we developed a tumor control dose-response model handling the multiple modes of failure for standard-risk MB (Study V). The aim was to provide a mathematical framework for estimating the effect on tumor control when varying the radiation treatment, either through varying the fractionation or the dose prescription. Ideally, such a tumor control model would be based on large randomized trials comparing several different radiotherapy dose levels. As this is not available for standard-risk MB we decided to pool all the available published outcome data from trials and prospective cohort studies, for patients treated with definitive radiotherapy and chemotherapy.

Given that recurrences can occur either in the primary boost site (posterior fossa) or in the craniospinal elective volume, the possibility of multiple modes of failure has to be considered. Pattern of failure data needs to be extracted, where treatment failures are recorded as occurring either in the boost site, the elective site or as synchronous failures in both sites. Mathematically, this can be described by the following failure frequencies:

\[
\rho_{\text{elective}}^{\text{recorded}}, \quad \rho_{\text{boost}}^{\text{recorded}}, \quad \rho_{\text{sync}}^{\text{recorded}}
\]

Clinically, truly synchronous failures are very unlikely, if the different failure modes are independent, and recorded synchronous failures could be a result of a tumor recurrence appearing in one site and spreading to the other before the patient is diagnosed at a follow-up examination. Handling synchronous failures can be tricky and assuming statistical independence between the two failure modes considerably underestimates the number of synchronous failures compared to published data, cf. Table 1 in Study V.
The mathematical derivation for obtaining the true failure frequencies is given in the methods section of Study V although the idea is to derive the \( \rho_{\text{sync}} \) and then obtaining the true failure rates in the boost and elective site. This is however dependent on the assumption of statistical independence \( \rho_{\text{elective}} \cdot \rho_{\text{boost}} = \rho_{\text{sync}} \) which may be or may not be correct. We performed a sensitivity analysis to assess the impact of this assumption on the model. This was done by introducing a parameter \( \eta \) so that:

\[
\rho_{\text{elective}} \cdot \rho_{\text{boost}} = \eta \cdot \rho_{\text{sync}}, \text{ where } \eta \leq 1 \tag{Eq. 5}
\]

and finding the lowest value of \( \eta \) that satisfies the condition \( \rho_{\text{true}} \leq \rho_{\text{simu}} \) when evaluated against the available clinical pattern of failure data. This would re-distribute more failures as being truly synchronous and thus scored as both elective and boost site failures, rather than boost site failures subsequently spreading to the elective volume. Fitting dose-response models for the boost and elective site to the clinical pattern of failure data resulted in the models shown in Figure 4. For the elective site, we show the models given the assumption of statistical independence and for the sensitivity analysis which yielded \( \eta = 0.72 \). As seen here there was no noticeable difference on the elective tumor control model from relaxing the assumption of statistical independence.
Figure 5. Resulting tumor control dose-response models for the boost and elective craniospinal site. Vertical bars show standard deviations of data points and the blue dashed lines give the 95% confidence intervals of the models.

Several different types of tumor control models have been proposed for radiotherapy purposes and evaluating a treatment plan will depend on the different underlying assumptions. There are two principally different model types; mechanistic models and phenomenological models. A mechanistic model is based on the interpretation of the underlying biological mechanisms, which in tumor control modelling can be clonogenic cell density, surviving fraction at a given dose, and tumor cell re-population over time. In this case the final tumor control model is often based on inputting pre-clinical experimental data to yield the tumor control probability (TCP) for a given tumor type. A phenomenological TCP model is fitted to the available clinical outcome data without considering the underlying biological mechanisms.

In our analysis we applied phenomenological TCP models for the elective and boost site, as shown in Figure 5, where the dose-response parameters of the model were fitted to the 5-year EFS at different doses. The full mathematical derivation and underlying assumptions is given in the methods in Study V. We estimated a rather shallow dose-response model for the elective craniospinal site for patients treated with radiotherapy and chemotherapy.
This is in contrast to treatment with radiotherapy alone, supporting the hypothesis that the shallow elective volume dose-response in standard-risk MB can be attributed to the effectiveness of chemotherapy in handling sub-clinical disease. This would provide great potential for reducing the risk of late complications by lowering the CSI dose even further. Lowering the CSI dose should, however, be done with caution and in a controlled setting to evaluate the efficacy. The large confidence intervals for CSI doses lower than 18 Gy, which are attributed the lack of data in this dose range, are especially concerning since lowering the CSI dose may result in high rates of elective site recurrences. The results of the ongoing ACNS 0331 trial randomizing 23.4 Gy CSI vs. 18 Gy CSI should provide further insight into whether the estimated shallow dose-response holds up down to this dose.

Using time-to-failure data from two large MB trials, we used our TCP model to estimate the freedom from progression (FFP) for standard-risk patients (detailed description in Study V). A useful application of this FFP model is to run hypothesis-generating in silico simulations of altered treatment regimens. It can also be used to estimate the required sample size for trials comparing different dose prescriptions. Comparing the two treatment arms in the SIOP PNET-4 trial using our FFP model suggests that the hyperfractionated arm may be slightly more effective, although a much larger sample size would be required to show a statistically significant difference.

**Simultaneous optimization of tumor control and toxicity**

To truly evaluate whether one cancer treatment regimen is better than another, this has to be measured through the efficacy of controlling the tumor as well as the risk of developing treatment-induced complications. Comparing these different measures is far from trivial although this is the essence of radiation oncology decision making, which physicians are faced with on a daily basis in the clinic. A method for simultaneously estimating the chance of tumor control and the risk of treatment toxicity in a combined measure was
proposed already in the early stage of the IMRT era by Källman et al.\textsuperscript{103} They derived a measure for complication-free tumor control which they denoted $P_+$:

$$P_+ = P_B - P_B \cap P_I$$  \hspace{1cm} (Eq. 6)

where $P_B$ denotes the probability of benefit (tumor control) and $P_I$ the probability of injury. They even provided the mathematical framework for an optimization algorithm focused on maximizing the $P_+$ in a treatment planning system. A similar suggestion of including models for healthy tissue complications and TCP models in an inverse planning algorithm was more recently proposed by Nahum and Uzan.\textsuperscript{104} They also suggested including the effect of functional imaging investigations and perhaps even biological markers for individual patients.

Both of these approaches are quite attractive as they would allow for simultaneous optimization of both tumor control and toxicity. However, as was also pointed out by Källman, the chance of tumor control cannot directly be compared to the risk of inducing a complication. In some cases a certain level of toxicity may be deemed acceptable if controlling the primary tumor and the difference in time-to-event between recurrence and a late complication may be several years or decades.

The focus of Study VI was to expand the methods for LYL estimation to also enable quantification of the LYL attributable to not controlling the primary disease, thus obtaining a common scale measure of tumor control and risk of toxicity. The necessary components for estimating the LYL due to the primary disease and late complications are:

- A dose-response model for estimating the time to disease progression
- Survival data for relapsing patients
- Dose-response models for estimating the risk of late complications
- Long-term survival data (\textgreater{} 10 years) for the studied patient group

23
Other than the data already included in the LYL estimation in Study II, we used our FFP model for standard-risk MB (Study V) and survival data from a recent report on relapsing MB patients receiving salvage treatment.\textsuperscript{105}

Details regarding the mathematical implementation are given in Study VI although the general idea is to estimate the disease-specific survival within a certain time after treatment. This is calculated via the risk of dying, conditional on having relapsed within that time frame. To this end, the probability of relapse can be extracted from the FFP model and corresponding hazard of relapse. The implementation of a FFP model in the LYL framework gives us two things; the LYL attributable to death from the primary disease and the survival function used to account for competing risks of dying when estimating the LYL due to late complications.

We applied this mathematical framework to compare different risk-adapted treatment approaches with the current standard treatment, details of which are presented in Study VI. In Figure 6, the results of two specific risk-adapted treatment strategies are compared:

1. Lowering the prescribed dose to the part of the craniospinal target encompassing the 1st to 10th thoracic vertebra (Th1-Th10) using 3D CRT; the rationale being that this region would include the most critical structures for the current LYL measure (heart, lungs, breasts)
2. A homogeneous dose of 36 Gy with proton therapy, which enables healthy tissue sparing without reducing the dose to the target volume.
Figure 6. Difference in total number of life years lost (LYL) between the standard dose prescription and two risk-adapted strategies. Vertical bars show the 95% confidence intervals of the LYL differences. The corresponding dose distributions are shown in the right hand panel.

Our hypothesis for the risk-adapted Th1-Th10 strategy was that if lowering the dose to a small part of the whole target volume that is close to critical structures, the small drop in TCP may be compensated by a reduced risk of severe late complications. Although this yielded a net reduction in LYL compared to the standard treatment, the total number of LYL was clearly dominated by the LYL attributable to the primary disease in this young patient group. For a uniform CSI dose of 24 Gy, the estimated LYL due to the primary disease were about 14 years, compared to about 1 year from late complications. Young age at diagnosis and poor survival for relapsing patients are likely the reasons for the large number of LYL attributable to the primary disease. This shows the importance of not reducing the high survival rates for standard-risk MB when attempting to reduce the risk of treatment complications. The results of the presented risk-adaption strategies should be interpreted with caution given the underlying uncertainties and assumptions in the models, which are discussed in Study VI.
Special applications of statistical methods in this thesis

When comparing long-term risk estimates, either between different treatment plans or treatment modalities, the main contribution to the statistical uncertainty will likely come from the dose-response parameters of the respective radiobiological model. Whether the dose-response models are linear, logistic or based on other mathematical relations, the uncertainty of the model can be estimated as long as the uncertainty of the parameters is reported, usually as the 95% confidence interval (CI). When calculating combination metrics such as the LYL e.g., there are uncertainty contributions from several different sources that go into the combined measure. That means it can be difficult to analytically estimate how much the uncertainty from each of the different dose-response models affects the final uncertainty in LYL.

In this thesis, we applied numerical Monte Carlo (MC) methods to incorporate uncertainty from several different sources. This was done by randomly sampling distributions matching the point estimate and corresponding 95% CI of the different dose-response parameters. We then ran the full calculation algorithm for estimating the LYL once for the point estimate of the parameters, and once for every random sample of the parameters. This leaves us with a point estimate and a distribution of randomly sampled LYL values, for each patient included in the analysis, with 95% CIs that represents the combined uncertainty of all models included in the calculation.

To estimate whether the difference between two treatments is statistically significant, we use the MC sampled LYL distributions to make pairwise comparisons of the LYL between the two treatments, for each of the random samples. This results in a distribution of LYL differences for each patient, as shown in the upper panel of Figure 7. The LYL difference for the whole patient group, including 95% CI, is calculated as the mean between all patients, weighted by the inverse variance. The difference is statistically significant at the 5% level if the low end of the 95% CI does not cross zero.
Figure 7. Example of estimating the 95% confidence interval (CI) of the difference in life years lost (LYL) between two different treatment modalities. The upper panel shows the Monte Carlo sampled LYL distributions for 10 patients and the bottom panel shows how bootstrapping is applied to yield the final mean estimate and CI. Reprinted from Brodin et al.106 with permission from Cancer.

When the number of patients is quite small, which is the case for most treatment planning comparisons, the variance in the studied cohort may not fully represent the variance for a larger group of patients. This would result in CIs that are too narrow to generally represent the studied patient group. To somewhat account for this, we applied a bootstrapping procedure when estimating the final LYL difference. By drawing random samples with replacement from the group of patients, we end up with different inverse variance
weighted LYL difference distributions for each sample; depending on how many times the
different patients were drawn. For each bootstrap sample, a LYL difference estimate is
randomly drawn from the corresponding distribution. The final bootstrapped LYL
difference with 95% CI is then taken as the mean and 2.5 – 97.5 percentile of the
randomly drawn samples, as illustrated in the bottom panel of Figure 7.

This methodology is not limited to estimating risk differences as in this example but can
be used to estimate odds ratios or hazard ratios between different interventions.

Summary and conclusions

Although MB patients are at risk of many different late complications we showed that
mathematical dose-response models can be used to estimate the risk of several different
complications and compare them between different treatment options. Using a LYL
measure, the various complications risks can be compared directly on a common scale. We
used this to illustrate the trade-off between irradiating a large volume of the thorax to a
low dose with craniospinal VMAT compared a high dose to the heart with craniospinal 3D
CRT. Here, proton therapy has a large advantage in the estimated number of LYL
attributed to the ability to limit the dose to the ventral part of the thorax.

We also showed that the ability to limit the dose to the neurogenic niches, especially the
hippocampus, is estimated to considerably reduce the risk of long-term neurocognitive
decline. Here, the inversely-planned radiotherapy techniques can reduce the hippocampal
dose; both from the cranial irradiation part of the CSI and the high-dose boost treatment.
Spot-scanned proton therapy performs significantly better than IMRT and VMAT in a
hippocampal sparing setting although the availability this treatment is currently sparse.
There is also an effect on estimated risk of cognitive decline from varying the treatment
margin applied in the high-dose boost, although the effect of applying hippocampal
sparing was greater than that of reducing the boost margin.
We developed a tumor control dose-response model for standard-risk MB where the dose-response for the elective CSI was estimated to be rather shallow when treatment also includes maintenance chemotherapy. Although very uncertain at CSI doses below 18 Gy the model suggests that limiting the dose to small areas of the craniospinal axis only yields a small drop in TCP. Using models of risk estimation with our tumor control model we estimated and compared the LYL from both the primary disease and late complications on a common scale. For these young patients, the LYL attributable to the primary disease dominated the total number of LYL, compared to the contribution from late complications. We showed, however, that it is possible to optimize treatment efficacy and toxicity estimated in a single LYL measure. A useful extension of the advanced dose-effect models applied here could be to run hypothesis-generating in silico trials or to make power calculations for upcoming investigations.

The general conclusions of this thesis are that the estimated risk of late complications in MB survivors can be considerably reduced either through advanced optimization of modern photon treatment or preferably by using spot-scanned proton therapy. A LYL-based optimization of proton treatment plans, including hippocampal sparing, shows great promise for reducing the risk of secondary cancers, cardiac complications and neurocognitive impairment in MB patients.

**Future perspectives in medulloblastoma research**

Research in pediatric MB is on the verge of a great step into further treatment individualization and considerably improved risk-stratification. In recent years it has been discovered that MB comprises four distinct molecular subgroups.\textsuperscript{107} Interestingly, when patients are stratified according to molecular classification, this is far more predictive of treatment outcome compared to risk-classification based on clinical variables.\textsuperscript{108} Two subgroups are especially interesting; the WNT subgroup with excellent long-term survival and the Group C subgroup, expressing MYC protein amplification, which has a very poor prognosis.\textsuperscript{107, 109, 110} This has prompted research efforts not only into targeted therapies for
the different groups but also suggestions of new risk-stratification for currently available radio- and chemotherapy (cf. Figure 5 in Ellison\textsuperscript{109}). An updated staging system is currently being implemented and will likely divide MB patients into a low-risk, an intermediate-risk and a high-risk group, based on clinical variables, histology and tumor biology.\textsuperscript{108, 111}

The upcoming SIOP PNET-5 trial will investigate the efficacy of less aggressive treatment in the low-risk patient group and similarly we can likely expect upcoming trials of increased treatment efforts in the group with the worst prognosis. Stratifying patients according to biological markers and performing pattern of failure analyses on the recently completed large MB trials could provide vital information for tailoring new risk-adapted treatment regimens. This should be done in a competing risk setting to properly account for the many severe complications attributable to the aggressive therapy. Access to risk-stratified pattern of failure data would allow for an extension of the ideas presented in Study III, IV and VI where sparing the hippocampus or lowering the target dose close to critical healthy structures may be best pursued in the low-risk group, given the low risk of tumor recurrence. In general, MB research could benefit greatly from a detailed pattern of failure analysis, where the precise anatomical positions of recurrences are recorded along with the radiation dose delivered to that point. This along with stratification according to molecular subgroup would really allow for individualized treatment adaptation.

Proton therapy has currently been used in MB treatment with the aim of reducing the risk of late complications. Proton therapy can potentially also be used to boost selected target areas to higher doses than what is possible with photons, if these are located close to dose-limiting critical structures. This can be combined with e.g. the great hippocampal sparing potential of proton therapy to further optimize the therapeutic ratio. Regardless of whether proton or photon therapy is considered, the ultimate goal of the radiobiological models and the all-course mortality LYL estimation presented in this thesis would be to implement them into a treatment planning system. Optimization driven directly by minimizing e.g. the number of QALYs lost could be used as a complement to optimizing the dose distribution within the patient. This would require a more complete set of dose-response models so that all important treatment complications can be considered. Changing the penalty function in the inverse treatment planning system so that the
optimization is driven by QALYs lost rather than absorbed dose should, however, be fairly straightforward. The biggest challenge lies within the uncertainty in long-term risk estimates and translating risks from older radiotherapy modalities to the advanced treatment options available today.

A future implementation of risk models and LYL measures as well as how individual histology and biology affects treatment efficacy, into treatment planning routine would indeed provide clinicians with a powerful decision making tool focused on highly individualized MB treatment.
Reference List


(62) Mu X, Bjork-Erikksson T, Nill S, et al. Does electron and proton therapy reduce the risk of radiation induced cancer after spinal irradiation for childhood
medulloblastoma? A comparative treatment planning study. Acta Oncol

(63) Ng AK, Kenney LB, Gilbert ES, Travis LB. Secondary malignancies across the age

(64) Mackie TR, Holmes T, Swerdloff S, et al. Tomotherapy: a new concept for the

(65) Yu CX. Intensity-modulated arc therapy with dynamic multileaf collimation: an

(66) Brodin NP, Munck Af RP, Aznar MC, et al. Radiobiological risk estimates of
adverse events and secondary cancer for proton and photon radiation therapy of


(68) Deutsch M, Thomas PR, Krischer J, et al. Results of a prospective randomized trial
comparing standard dose neuraxis irradiation (3,600 cGy/20) with reduced
neuraxis irradiation (2,340 cGy/13) in patients with low-stage medulloblastoma. A
Combined Children's Cancer Group-Pediatric Oncology Group Study. Pediatr

medulloblastomas with reduced-dose craniospinal radiation therapy and adjuvant
Jul;17(7):2127-36.

(70) Thomas PR, Deutsch M, Kepner JL, et al. Low-stage medulloblastoma: final
analysis of trial comparing standard-dose with reduced-dose neuraxis irradiation. J


(77) Kortmann RD, Kuhl J, Timmermann B, et al. Postoperative neoadjuvant chemotherapy before radiotherapy as compared to immediate radiotherapy followed by maintenance chemotherapy in the treatment of medulloblastoma in


PAPERS
STUDY I
ORIGINAL ARTICLE

Radiobiological risk estimates of adverse events and secondary cancer for proton and photon radiation therapy of pediatric medulloblastoma

N. PATRIK BRODIN1,2, PER MUNCK AF ROSENSCHÖLD1, MARIANNE C. AZNAR1, ANNE KIIL-BERTHELESEN1,3, IVAN R. VOGELIUS1, PER NILSSON3, BIRGITTA LANNERING5 & THOMAS BJÖRK-ERIKSSON4

1Radiation Medicine Research Center, Department of Radiation Oncology, Rigshospitalet, University of Copenhagen, Denmark, 2Niels Bohr Institute, Faculty of Sciences, University of Copenhagen, Denmark, 3Department of Clinical Physiology and Nuclear Medicine, Centre of Diagnostic Investigations, Rigshospitalet, University of Copenhagen, Denmark, 4Department of Oncology, Skåne University Hospital and Lund University, Lund, Sweden, and 5Department of Paediatric Oncology, The Queen Silvia Children’s Hospital, Gothenburg, Sweden

Abstract

Introduction. The aim of this model study was to estimate and compare the risk of radiation-induced adverse late effects in pediatric patients with medulloblastoma (MB) treated with either three-dimensional conformal radiotherapy (3D CRT), inversely-optimized arc therapy (RapidArc® (RA)) or spot-scanned intensity-modulated proton therapy (IMPT). The aim was also to find dose-volume toxicity parameters relevant to children undergoing RT to be used in the inverse planning of RA and IMPT, and to use in the risk estimations. Material and methods. Treatment plans were created for all three techniques on 10 pediatric patients that have been treated with craniospinal irradiation (CSI) at our institution in 2007–2009. Plans were generated for two prescription CSI doses, 23.4 Gy and 36 Gy. Risk estimates were based on childhood cancer survivor data when available and secondary cancer (SC) risks were estimated as a function of age at exposure and attained age according to the organ-equivalent dose (OED) concept. Results. Estimates of SC risk was higher for the RA plans and differentiable from the estimates for 3D CRT at attained ages above 40 years. The risk of developing heart failure, hearing loss, hypothyroidism and xerostomia was highest for the 3D CRT plans. The risks of all adverse effects were estimated as lowest for the IMPT plans, even when including secondary neutron (SN) irradiation with high values of the neutron radiation weighting factors (WRn). Conclusions. When comparing RA and 3D CRT treatment for pediatric MB it is a matter of comparing higher SC risk against higher risks of non-cancer adverse events. Considering time until onset of the different complications is necessary to fully assess patient benefit in such a comparison. The IMPT plans, including SN dose contribution, compared favorably to the photon techniques in terms of all radiobiological risk estimates.

Cancers of the central nervous system (CNS) are, second to lymphomas and leukemias, the most common types of childhood cancers corresponding to 20–30% of all pediatric cases [1,2]. The most common malignant CNS tumor in children is medulloblastoma (MB) which is a primitive neuro-ectodermal tumor located in the posterior cranial fossa. MBs are characterized by a relatively high rate of spinal metastases at the time of diagnosis [3,4]. The overall five-year survival for standard risk MB (patients with primary tumor in the posterior fossa but without confirmed spinal metastases and with negative cerebrospinal fluid cytology) is 75–85% for children treated with modern day multimodality regimens [5,6]. Treatment consists of surgery followed by post-operative multi-agent chemotherapy and craniospinal irradiation (CSI). Survival rates are relatively high but there are substantial long-term side effects related to radio-chemotherapy of pediatric patients; including loss of hearing or vision,
neurocognitive deficits, gonadal dysfunction, cardiopulmonary impairment and endocrine effects [7–16]. The concern regarding therapy-induced secondary cancer (SC) has increased in recent years and is especially relevant for childhood cancer survivors. It has been shown that young age at treatment is associated with an increase in risk for development of most SCs [17,18]. Several approaches have been suggested for estimating the risk of SC induction via different models with dose-responses relating to either linear, linear-exponential or linear-quadratic behavior [19–23]. When comparing results of SC risk estimates from different studies it is vital to consider whether it concerns the risk of contracting a SC or the risk of dying from that SC as this endpoint may vary between studies. It is also important to stratify if the estimates concern lifetime risk or the risk up to a certain attained age as these estimates will differ. Estimating the SC risk on an organ-specific basis would provide the option of assessing whether limiting the dose to the organs most responsible for SC induction can lower the total SC risk.

The prescribed CSI dose has been successfully decreased from 36 Gy to 23.4 Gy in standard-risk MB patients without compromising survival rates [24]. Provided that the disease is properly staged as either standard or high risk, the treatment regimen can be stratified accordingly. Common to both standard- and high-risk patients is that the CSI treatment is followed by a boost to the primary tumor site (posterior cranial fossa) up to 54–55 Gy [5]. Since radiotherapy can be given with varying prescribed CSI doses it is warranted to consider the impact of this variation in estimations of treatment-related long-term complications. There is lack of general consensus on the lateral extent of the spinal part of the target volume. Including the entire vertebral column is a viable option as this prevents asymmetric spinal growth. This is based on the findings that bone growth is inhibited at absorbed doses above approximately 20 Gy [25].

Today the standard technique for CSI is three-dimensional conformal radiotherapy (3D CRT). This technique results in irradiation of a large volume of normal tissue to high doses. Several studies have investigated the feasibility of CSI with intensity-modulated radiotherapy (IMRT) or proton therapy and the estimated effect on SC risk [21,26–28].

This study was aimed at estimating and comparing SC risk and risks of several other non-cancer long-term complications between 3D CRT, rotational IMRT and spot-scanned intensity-modulated proton therapy (IMPT). The aim was also to use the inverse optimization tools in IMRT and IMPT to purposely limit the dose to most organs at risk (OARs), as an attempt at radiobiological treatment optimization for pediatric MB. Using the full potential of the different treatment techniques in this way allows for a realistic comparison of modalities. The treatment plans and risk estimations were performed for varying CSI prescription doses and spinal target volumes in order to gain some perspective into the effect of these variations on long-term complications.

**Material and methods**

**Patient material and treatment planning**

Ten pediatric patients, four males and six females, ages 4 to 15 years (mean age 8 years), were included in this study. They had all received CSI for MB at the Copenhagen University Hospital in 2007–2009. The craniospinal clinical target volume (CTV) was defined as comprised of a cranial CTV consisting of the whole brain and a spinal CTV defined as the spinal canal extending caudally to the S2-S3 junction. Target volume definitions were based on computed tomography (CT) scans. The craniospinal planning target volume (PTV) consisted of the cranial CTV with a 5 mm isotropic margin and the spinal CTV with a 7 mm isotropic margin. An alternative PTV was created with the spinal part consisting of the entire vertebral column with 3 mm lateral margins. This alternative PTV was not considered for the main analysis but rather to test how different definitions of spinal target volume would affect the risk estimates. To facilitate treatment planning comparison, prescription doses were set as 23.4 Gy and 36 Gy to the craniospinal PTV. The fractionation scheme was chosen as 1.8 Gy/fraction for all prescription levels and each prescription regimen was followed by boost treatment to a total dose of 54 Gy to the boost PTV comprised of the posterior cranial fossa with a 5 mm isotropic margin.

All treatment plans were generated using the Eclipse treatment planning system version 8.9 (Varian Medical Systems, Palo Alto, CA). The different techniques were normalized to all have the same mean target dose as the 3D CRT plans, in order to facilitate objective plan comparison. The 3D CRT plans were created with two lateral opposed cranial fields. The caudal edge of these fields was positioned just above the cranial edge of the patient’s shoulders in order to protect the thyroid gland. The eyes were shielded by MLC-leaves but only to the extent that it did not compromise PTV coverage. Additionally, a spinal posterio-anterior field encompassing the spinal part of the PTV was applied. The junction between the cranial and spinal fields was
moved 1.0 cm caudally once in the 23.4 Gy prescription regimen and twice in the 36 Gy regimen to limit over or under dosage due to possible set-up variations in a clinical treatment setting. In order to assess the importance of the lateral extent of the spinal target volume in 3D CRT, multiple plans were created with varying spinal field widths. For the different plans the multi-leaf-collimator (MLC) leaves were fitted to the spinal PTV with 0.5 cm, 0.75 cm, 1.0 cm and 1.5 cm lateral margins, respectively. When reviewing the treatment records of the patients, the 1.0 cm leaf-to-PTV margin corresponded best to the set-up in which most of the patients had been treated. This margin was chosen when comparing the 3D CRT plans to the other techniques. The boost plans were created using four fields, two opposing fields from the right and left sides along with two wedged fields incident from a more posterior direction.

Rotational IMRT plans were created using the RapidArc® (RA) technique (Varian Medical Systems, Palo Alto, CA). RA plans consisted of a cranial arc covering the whole brain and the cranial part of the spinal canal down to the C2 or C3 vertebra. The rest of the spinal canal was covered by a spinal arc delivering radiation from a posterior 140 degree arc only to avoid unnecessary irradiation through the arms and ventral parts of the patient. There was substantial overlap between the cranial and spinal arcs to facilitate an intensity-modulated junction that was less sensitive to set-up error than a sharp junction between field edges. Two spinal arcs were used in the case where the cranio-caudal extent of the spinal canal exceeded the span of a single arc. The boost plans consisted of a 220 degree arc covering the boost PTV without irradiating through the anterior parts of the head.

IMPT plans were generated with machine data corresponding to a PT2 Varian proton therapy system utilizing active spot-scanning. The plans were generated using three fields incident from the posterior direction. A range shifter can be introduced when necessary, in order to reach both shallow and deeply situated structures in the body. The range shifter effectively lowers the span of the achievable nominal energy, which is 70–250 MeV for this machine, and thus decreasing the depth to which the protons reach. The IMPT plans consisted of one field with a range shifter covering most of the spinal canal and two fields, of which one had a range shifter, covering the brain and the cranial part of the spinal canal. The junction between the IMPT fields was set with an overlap and modulated between fields as for the RA technique. The boost plans consisted of a single posterior field with a range shifter to ensure full coverage of the posterior fossa. Figure 1 illustrates the difference in dose distribution relating to the different treatment techniques.

Proton radiotherapy exposes the patient to secondary neutron (SN) irradiation which is not considered in the treatment planning system hence this extra dose contribution was added manually. This was done by using organ-specific neutron doses corresponding to a spot-scanned proton beam taken from the study by Newhauser et al. [26]. In that study the neutron dose contribution was simulated using Monte Carlo methods for a pediatric mathematical phantom receiving craniospinal proton irradiation. A limitation of using this data is that SNs are assumed to be generated within the patient only. This will underestimate the SN contribution, especially when a range shifter is applied as the beam passing through the range shifter will induce SNs. In order to account for possible uncertainty in the relative biological effectiveness (RBE) of SNs, neutron radiation weighting factors...
The inverse planning optimization for RA and IMPT was focused on lowering the risk of treatment-induced complications by limiting the dose to as many critical structures as possible. The actual optimization objectives were set with the aim of fulfilling the dose objectives proposed in Table I, when possible. The data shown in Table I are the results of an extensive literature search focused on finding dose tolerance levels relevant for pediatric MB patients. The dose objectives and the inter-organ priority were based on the references given in Table I and the preferences of experienced physicians in pediatric- and general oncology (BL, TBE). The main difference between RA and IMPT optimization was the fact that most of the OARs did not receive any absorbed dose from the protons as the distal edge of a beam is very sharp.

Modeling of secondary cancer risk and non-cancer adverse effects

A concept for phenomenological modeling of SC risks is that of organ equivalent dose (OED) which was developed by Schneider et al. in 2005 [22]. This provides a single measure of an OED in Gy which represents the often heterogeneous dose distribution of a dose-volume histogram (DVH). This value represents the uniform dose that would yield the same SC risk as an organ as the inhomogeneous dose distribution described in the DVH. An alternative way of estimating OED was presented by Schneider et al. in 2008 based on the combined SC data of Hodgkin’s disease patients having undergone radiotherapy and that of the atomic bomb survivors extrapolated to higher doses more relevant to radiotherapy [30]. The authors fitted this data to OED models with a linear, linear-exponential (bell-shaped) and plateau dose-response, although the linear model did not provide a good fit to data. There is evidence from large clinical studies of SC that does not support a decrease in risk at high radiotherapy doses [31,32]; the plateau-model thus appears to be a more appropriate selection/model. Even with appropriate models the uncertainties in SC risk estimates are considerable. When faced with clinical decisions one needs to consider the differences, for example between modern treatment and that given several decades ago and the possible variations in genetic cancer susceptibility between the atomic bomb survivors and modern day cancer patients.

The OED for the plateau-model was calculated as

$$OED_{org} = \frac{1}{N} \sum_{i=1}^{N} 1 - e^{-\delta_{org,i}/\delta_{org}}$$  \hspace{1cm} (1)$$

where $\delta_{org}$ is the organ-specific dose-response parameter derived from fitting the model to the Hodgkin’s cohort data. $D_i$ is the dose to bin $i$ in the DVH and $N$ is the total number of dose bins. To test the sensitivity of the SC risk on the choice of model, estimates were also made with the bell shaped and linear models from Schneider et al. [30] as a comparison. The excess absolute risk (EAR) of SC was estimated as a linear function of the OED adjusted for population-specific variables according to Equation 2.

$$EAR(D,e,a,s) = \beta OED(a,s)$$  \hspace{1cm} (2)$$

The parameter $\mu$ represents the population-specific risk modification and is a function of age at exposure, $e$, attained age, $a$ and gender, $s$, and $\beta$ describes the initial slope of the dose-response curve [30].

$$\mu(e,a,s) = \exp \left[ \gamma_e (e - 37) + \gamma_a \ln \left( \frac{a}{46} \right) \right] (1+s)$$  \hspace{1cm} (3)$$

The constants $\gamma_e$ and $\gamma_a$ are related to attained age and age at exposure and were derived from fitting the model to the original data. The cumulative EAR of developing a SC was then estimated to a given attained age by taking into account competing risks of death according to Equation 4.

$$C_{EAR}(a) = \sum_{i=e+1}^{\infty} EAR(D_i,e,a,s) \cdot \frac{S(i)}{S(e)}$$  \hspace{1cm} (4)$$

The ratio $S(i)/S(e)$ represent the gender-specific conditional probability of a person alive at age $e$ to reach age $i$. The survival functions were taken from Koller et al. [33].

A limitation of the model by Schneider et al. is that no organ-specific model parameters were derived in the original publication [30]. The dose-response curve relates to the induction of any solid secondary cancer, thus representing the mean dose-response for all secondary cancers studied. In order to calculate the organ-specific risk contribution, the dose-response was assumed to be valid for all studied organs.

To take the risk contribution from each of the relevant organs into account, a weighted mean OED was calculated for all organs for which incidence of radiation-induced cancers was published in the Life Span Study (LSS) [34]. The weighted mean OED was calculated based on the data provided in Table II according to Equation 5.
Table I. Organ-specific dose objectives relating to children treated with radiation therapy.

<table>
<thead>
<tr>
<th>Irradiated organ</th>
<th>Endpoint</th>
<th>Patient data</th>
<th>Chemo</th>
<th>NTCP-model</th>
<th>Priority</th>
<th>Dose objectives</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone Marrow</td>
<td>Cataract</td>
<td>Children</td>
<td>N/A</td>
<td>N/A</td>
<td>2</td>
<td>D_{mean,unirrad} &lt; 10 Gy</td>
<td>Chang et al. (2002)</td>
</tr>
<tr>
<td></td>
<td>Dry eyes</td>
<td>Children</td>
<td>N/A</td>
<td>N/A</td>
<td>2</td>
<td>D_{mean} &lt; 20 Gy or BED (mean eye dose) &lt; 40 Gy, et</td>
<td>Kal et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Blindness</td>
<td>Children</td>
<td>N/A</td>
<td>N/A</td>
<td>2</td>
<td>D_{mean} &lt; 10 Gy</td>
<td>Whelan et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Cataract</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>1</td>
<td>D_{mean,heart} &lt; 2.5 Gy</td>
<td>Goldner et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>Radiation Induced Liver Disease</td>
<td>Adults</td>
<td>No</td>
<td>LKB</td>
<td>1</td>
<td>D_{mean,unirrad} &lt; 10 Gy</td>
<td>Dawson et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>Radiation Induced Liver Disease</td>
<td>Adults</td>
<td>No</td>
<td>N/A</td>
<td>1</td>
<td>D_{mean,unirrad} &lt; 10 Gy</td>
<td>Xu et al. (2006)</td>
</tr>
<tr>
<td>Neurocognitive effects</td>
<td>IQ decrease</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>2</td>
<td>D_{mean,unirrad} &lt; 25 Gy</td>
<td>Merchant et al. (2008)</td>
</tr>
<tr>
<td>Parotids</td>
<td>Xerostomia (75% flow rate red.)</td>
<td>Adults</td>
<td>N/A</td>
<td>LKB</td>
<td>3</td>
<td>EUD_{parotid} &lt; 14 Gy</td>
<td>Houweling et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Xerostomia (75% flow rate red.)</td>
<td>Adults</td>
<td>N/A</td>
<td>LKB</td>
<td>3</td>
<td>EUD_{parotid} &lt; 14 Gy</td>
<td>Roesink et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Xerostomia (50% flow rate red.)</td>
<td>Adults</td>
<td>N/A</td>
<td>N/A</td>
<td>3</td>
<td>EUD_{parotid} &lt; 14 Gy</td>
<td>Bussel et al. (2004)</td>
</tr>
<tr>
<td>Endocrine system</td>
<td>Hypothyroidism</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>5</td>
<td>D_{mean,thyroid} &lt; 6 Gy</td>
<td>Ricardi et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Hypothyroidism</td>
<td>Adults</td>
<td>Y &amp; N</td>
<td>N/A</td>
<td>4</td>
<td>D_{mean,unirrad} &lt; 20 Gy</td>
<td>Bhatia et al. (1996)</td>
</tr>
<tr>
<td></td>
<td>Neurologic sequelae</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>-</td>
<td>-</td>
<td>Xu et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>Sitting height</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>-</td>
<td>-</td>
<td>Rose et al. (2005)</td>
</tr>
<tr>
<td>Gynecological</td>
<td>Infertility</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>3</td>
<td>D_{mean,infertility} &lt; 40 Gy</td>
<td>Green et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Infertility</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>3</td>
<td>D_{mean,infertility} &lt; 50 Gy</td>
<td>Green et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Radiation induced menopause</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>3</td>
<td>D_{mean,infertility} &lt; 30 Gy</td>
<td>Chiarelli et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>Infertility</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>3</td>
<td>D_{mean,infertility} &lt; 10 Gy</td>
<td>SFRO (2008)</td>
</tr>
<tr>
<td>Kidneys</td>
<td>Chronic renal dysfunction</td>
<td>Children</td>
<td>N/A</td>
<td>N/A</td>
<td>1</td>
<td>BED_{mean} &lt; 16 Gy</td>
<td>Cal et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Renal toxicity</td>
<td>Children</td>
<td>N/A</td>
<td>N/A</td>
<td>3</td>
<td>D_{mean,unirrad} &lt; 10 Gy</td>
<td>SFRO (2008)</td>
</tr>
<tr>
<td></td>
<td>Restrictive lung function</td>
<td>Children</td>
<td>No</td>
<td>N/A</td>
<td>1</td>
<td>D_{mean,unirrad} &lt; 10 Gy</td>
<td>Attard-Montalto et al. (1992)</td>
</tr>
<tr>
<td></td>
<td>Pulmonary diffusion capacity</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>1</td>
<td>D_{mean,unirrad} &lt; 10 Gy</td>
<td>Bossi et al. (1997)</td>
</tr>
<tr>
<td></td>
<td>Pulmonary function reduced</td>
<td>Children</td>
<td>No</td>
<td>N/A</td>
<td>1</td>
<td>D_{mean,unirrad} &lt; 10 Gy</td>
<td>Weiner et al. (2006)</td>
</tr>
<tr>
<td></td>
<td>Pneumonitis</td>
<td>Adults</td>
<td>Yes</td>
<td>Mean Lung Dose</td>
<td>1</td>
<td>D_{mean,unirrad} &lt; 10 Gy</td>
<td>Marks et al. (2010)</td>
</tr>
<tr>
<td>Hearing apparatus</td>
<td>Otoxicity</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>2</td>
<td>D_{mean,unirrad} &lt; 37 Gy</td>
<td>Miettinen et al. (1997)</td>
</tr>
<tr>
<td></td>
<td>Otoxicity</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>2</td>
<td>D_{mean,unirrad} &lt; 37 Gy</td>
<td>Huang et al. (2002)</td>
</tr>
<tr>
<td>Submandibular glands</td>
<td>Xerostomia (75% flow rate red.)</td>
<td>Adults</td>
<td>N/A</td>
<td>N/A</td>
<td>5</td>
<td>D_{mean,unirrad} &lt; 19 Gy</td>
<td>Mourtzakis-Kinch et al. (2008)</td>
</tr>
<tr>
<td>Chiasm</td>
<td>Blindness</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>2</td>
<td>D_{mean,unirrad} &lt; 52 Gy &amp; V_{100%} &lt; 100%</td>
<td>SFRO (2008)</td>
</tr>
<tr>
<td>Optic nerve</td>
<td>Blindness</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>2</td>
<td>D_{mean,unirrad} &lt; 50 Gy</td>
<td>SFRO (2008)</td>
</tr>
<tr>
<td>Skin</td>
<td>Telangiectasia/lipitation</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>5</td>
<td>D_{mean,unirrad} &lt; 35 Gy</td>
<td>SFRO (2008)</td>
</tr>
<tr>
<td>Brainstem</td>
<td>Necrosis</td>
<td>Children</td>
<td>Yes</td>
<td>N/A</td>
<td>1</td>
<td>D_{mean,unirrad} &lt; 54 Gy</td>
<td>SFRO (2008)</td>
</tr>
<tr>
<td>Esophagus</td>
<td>Esophageal RTOG Grade 1-4 tox</td>
<td>Adults</td>
<td>Y &amp; N</td>
<td>N/A</td>
<td>3</td>
<td>D_{mean,unirrad} &lt; 40 Gy</td>
<td>Ahn et al. (2005)</td>
</tr>
<tr>
<td>Larynx</td>
<td>Laryngeal edema</td>
<td>Adults</td>
<td>Y &amp; N</td>
<td>LKB-EUD</td>
<td>3</td>
<td>EU_{mean} &lt; 29 Gy</td>
<td>Raneil et al. (2009)</td>
</tr>
</tbody>
</table>

1Referenced publication is based on the gathered clinical experience of the Société Française de Radiothérapie Oncologique (SFRO), not on published clinical data, published as Guide des procedures de radiothérapie externe 2007 in 2008. EU_{D}, Equivalent Uniform Dose, further explanation given with corresponding reference.

Inter-organ priorities were based on:
1. Permanent, life threatening event with no possible medical substitution of organ function.
2. Permanent, non-life threatening event with inability to lead an independent life, medical substitution not possible.
3. Permanent effect with possibility of partial medical substitution.
4. Permanent effect with possibility of full medical substitution but with likely complications.
5. Permanent effect with possibility of complication free full substitution.
Radiobiological risk estimation in the treatment of pediatric medulloblastoma

3.35 2.29 – –

Leukemia

0.04 0.04 – –

Hodgkin’s disease

Multiple myeloma

0.26

A function of radiation dose [35]. We thus estimate the model could be found that well predicts the risk as estimating the risk of secondary leukemia but no excluded (shown in italics). An attempt was made at contracting a localized solid SC. Negative values in Table II are interpreted as zero and thus not contributing to the secondary cancer risk.

Table III. Radiation-related secondary cancer incidence from the LSS [34] and as adopted in this study. The incidence is given as average excess absolute risk due to external low-LET radiation as per 10 000 person years and Sieverts equivalent dose.

<table>
<thead>
<tr>
<th>Site/cancer type</th>
<th>Incidence (10^4 PYSv)^{-1}</th>
<th>Values adopted (10^4 PYSv)^{-1}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td>Esophagus</td>
<td>0.26</td>
<td>0.44</td>
</tr>
<tr>
<td>Stomach</td>
<td>2.61</td>
<td>5.86</td>
</tr>
<tr>
<td>Colon</td>
<td>2.66</td>
<td>1.01</td>
</tr>
<tr>
<td>Liver</td>
<td>0.033</td>
<td>0.005</td>
</tr>
<tr>
<td>Lungs</td>
<td>2.67</td>
<td>5.81</td>
</tr>
<tr>
<td>Bone and connective tissue</td>
<td>0.38</td>
<td>0.12</td>
</tr>
<tr>
<td>Skin</td>
<td>0.89</td>
<td>0.72</td>
</tr>
<tr>
<td>Female breast</td>
<td>–</td>
<td>6.80</td>
</tr>
<tr>
<td>Prostate</td>
<td>0.44</td>
<td>–</td>
</tr>
<tr>
<td>Urinary bladder</td>
<td>0.84</td>
<td>1.02</td>
</tr>
<tr>
<td>Brain and CNS</td>
<td>–0.21</td>
<td>0.43</td>
</tr>
<tr>
<td>Thyroid</td>
<td>0.87</td>
<td>2.32</td>
</tr>
<tr>
<td>Non-Hodgkin’s lymphoma</td>
<td>0.73</td>
<td>–0.20</td>
</tr>
<tr>
<td>Hodgkin’s disease</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>0.26</td>
<td>–0.08</td>
</tr>
<tr>
<td>Leukemia</td>
<td>3.35</td>
<td>2.29</td>
</tr>
</tbody>
</table>

As the ability of organ-specific SC risk estimation is dependent of knowing where a SC might occur, the cancer types where no localized anatomical site relating to radiation exposure could be specified were excluded (shown in italics). An attempt was made at estimating the risk of secondary leukemia but no model could be found that well predicts the risk as a function of radiation dose [35]. We thus estimate the risk of contracting a localized solid SC. Negative values in Table II are interpreted as zero and thus not contributing to the secondary cancer risk.

Modeling of non-cancer normal tissue complication probabilities (NTCP) was done according to the radiation dose-response functions presented in Table III. The basic rationale of NTCP modeling is to use the dose-response relationship of a complication derived from a patient cohort and apply it as a predictor of complication rate in other patients. This is used to define dose limits for OARs that should keep the risk of complications at an acceptable level. Although NTCP modeling can be a powerful tool in treatment planning, the model parameters are associated with considerable uncertainty. Relative comparison of pairs of treatment plans or modalities are however less sensitive to these uncertainties than absolute risk estimates.

Results

The estimations in Figure 2 show that the SC risk is highest for the RA treatments but the difference between the photon techniques does not become apparent until attained ages above 40 years, representing a latency period since treatment of about 30 years. The estimate of solid SC risk at an attained age corresponding to the average lifetime of a Danish person (78.5 years) was 45%, 56% and 7% for 3D CRT, RA and IMPT, respectively, for 23.4 Gy prescribed CSI dose. The corresponding risks for the 36 Gy CSI dose were 54%, 71% and 9%. These results were compared with the corresponding EARs estimated with a bell shape and a linear model as shown in Figure 3. It is shown that the choice of model does affect the absolute risk estimates especially for the RA plans. The absolute estimates depend strongly on the initial slope of the different dose-response curves. The EAR estimate at an attained age of 78.5 years was lower for RA than for 3D CRT.
results are given as including SN dose with values of the WR neutron probability of surviving to older ages is very low. For IMPT the 100 years, after this the cumulative risk levels out since the this patient group. The results are shown up to an attained age of the two different prescription CSI dose levels that are relevant for different treatment modalities studied. Results are given for secondary cancer risk as a function of attained age for the three Figure 2. Mean values for all 10 patients of cumulative solid according to the linear model, 33% vs. 36% for the 23.4 Gy CSI dose and 50% vs. 54% for the 36 Gy CSI dose. This is attributable to the fact that the mean doses to the SC critical structures were lower with RA and any linear model will thus favor RA over 3D CRT for these patients because of the direct relationship to mean dose.

It is shown that the excess absolute SC risk estimated for IMPT is affected by a large variation in the neutron RBE. The risk is however substantially lower with IMPT than for both photon techniques even with high values of WR_{neutron}.

Figure 4 shows the results of the NTCP estimations calculated according to Table III. The risk of inducing heart failure, xerostomia, hypothyroidism and ototoxicity was estimated as substantially higher for 3D CRT than for RA. The risk of pneumonitis was equivalent for the two photon techniques while the risk of inducing blindness was slightly higher for the RA plans. All NTCP estimates were lower with IMPT compared to the photon techniques even though the results in Figure 4 are presented for the IMPT_SNx5 plans. The risk of developing ototoxicity was shown to be higher for the lower CSI prescription dose for the RA plans. This was due to the boost plans with RA yielding a higher dose to the cochlea per prescription Gy than the main craniospinal plans. The estimates of chronic renal failure were not shown since the kidney doses were sufficiently low to not yield any considerable risk of renal failure; the highest estimate for any plan was below 2%. As NTCP estimates are subject to uncertainty, the absolute values should be considered somewhat cautiously with more emphasis on the relative comparison between modalities.

The risk of inducing heart failure was considerably higher for 3D CRT compared to the other two techniques. The effect on this endpoint of varying the spinal field widths was studied separately and shown in Figure 5. The risk of heart failure increases with spinal field width since the mean heart dose becomes larger. This shows that depending on the choice of field set-up in 3D CRT the risk of inducing heart failure is 2–2.5 times higher than that of RA.

Table IV represents target coverage parameters for the three different treatment techniques normalized to the same mean craniospinal PTV dose. The cranial target coverage was slightly better with 3D CRT compared to RA. The lack of conformity in the 3D CRT plans was shown by the high value of V_{107%} in the spinal CTV and the low scores in conformity index.

Discussion

The results of this study show the estimated radiobiological differences between 3D CRT, RA and IMPT in the treatment of pediatric MB. As expected the solid SC risk was higher for intensity-modulated photon treatment compared to 3D CRT. The EAR estimates are relatively high but we believe that this is the case for pediatric MB patients. Estimates are in reasonable accordance with the study by Mu et al.,
who also focused on pediatric MB. They estimated the lifetime risk of contracting a lethal SC to be 20% for 3D CRT and 30% for IMRT relating to a 23.4 Gy CSI prescription dose [21]. In the study by Miralbell et al. the yearly risk of SC induction was estimated for a 3-year-old MB patient to 0.76%, 0.43% and 0.05% for 3D CRT, IMRT and proton therapy, respectively [36]. This would correspond to respective lifetime risks of 55%, 31% and 4%, their estimates were based on a prescribed dose to the craniospinal axis of 36 Gy. The high risk related to 3D CRT in their study was likely attributed to the very young age of the patient as the treatment field would cover a large proportion of such a small body. A recent study estimated the long-term mortality in childhood cancer survivors based on the current follow-up of the childhood cancer survivor study (CCSS) cohort [37]. The authors showed that compared to other pediatric cases, MB patients were at high risk of SC mortality despite the fact that MB patients were the most prone to die from late recurrences.

Estimates of SC risk are in general subject to considerable uncertainty and often affected by the choice of model as shown in Figure 3, so the absolute estimates should be treated with caution. However, relative comparisons between treatment modalities can likely be more reliably considered. Treating children with IMRT, whether rotational or fixed-field,

Figure 4. Mean values for all 10 patients comparing long-term risks of inducing pneumonitis, heart failure, xerostomia, blindness, hypothyroidism and ototoxicity between the different treatment modalities.

Figure 5. Mean values of all 10 patients comparing the risk of heart failure for different spinal field widths in 3D CRT treatment.
Calculation details described elsewhere [45].

CI: Conformity Index. Closer to unity equals better conformity.

RapidArc techniques are more commonly used, the definition to occur. As modern, highly conformal, radiotherapy detriment and when the late effects are most likely to be considered as somewhat controversial.

should because of the underlying uncertainties still be considered as somewhat controversial.

The time since exposure aspect needs to be considered when trying to estimate actual patient detriment and when the late effects are most likely to occur. As modern, highly conformal, radiotherapy techniques are more commonly used, the definition of target volume becomes increasingly important. If, for example the entire vertebral column is included as PTV to receive at least 20 Gy we found that the SC risk was slightly increased for the RA plans and nearly doubled for the IMPT plans (data not shown). This results in weighting the risk of asymmetrical spinal growth, i.e. scoliosis, against that of an increased risk of SC. For such a comparison it is vital to consider at what ages the different complications are most likely to occur.

As for the risk estimations for non-cancer adverse effects, the results suggest that patients treated with 3D CRT are subject to substantially higher risks of serious complications such as heart failure and severe hearing loss compared to those treated with intensity-modulated therapy. The choice between the two photon techniques thus appears to be a choice between higher SC risk or higher risk of non-cancer adverse events. The basis for these risk comparisons lies in mathematical modeling and the inherent uncertainties need to be considered if the estimates are to guide clinical decision making. There is also the consideration of which technique provides the most robust treatment. A problem in 3D CRT treatment of MB is that it uses junctions between fields, risking over/under dosage due to uncertainties in treatment set-up. Intensity-modulated treatment would not suffer from the same problem since the field junctions can be modulated with intentional overlap between different fields. An error in treatment set-up would in this situation lead to only a small dose variation compared to the sharp junctions in 3D CRT.

The estimates of heart failure are based on patients that were treated with radiation therapy and anthracyclin chemotherapy, which also contributes to the risk of heart failure [38]. The radiotherapy related heart failure risk was about five times higher (data not shown) for 3D CRT compared to RA attributable to a large difference in mean heart dose (18.9 Gy vs. 7.3 Gy for the 36 Gy CSI dose). It is difficult to assess which risks are of greatest importance for the future of these patients. It might, for example be worse to have a 10% risk of developing heart failure before the age of 40 years than a 30% risk of contracting a SC before turning 60. Early results of comparing the effect of different long-term complications on a common scale have recently been shown and appear promising [39]. Estimating risks at old ages for people having survived a cancer in childhood is particularly difficult. This is primarily due to that cohorts in large studies like the CCSS [18] have not yet reached the ages where the general population incidence of cancer and heart complications becomes large. Increasing follow-up in, for example the CCSS will hopefully reveal whether the relative risks at older ages continue to be elevated, or whether they will decrease as the general population incidence increases.

In this work the optimization priority of normal tissue toxicity prevention was based solely on the severity of the different endpoints. The best treatment plans with respect to these risks would likely also include consideration of the incidence of the different complications. This would prevent the optimizer from trying to reduce a risk which is already negligible. We also attempted to optimize the RA plans with respect to SC risk by lowering the dose to the organs suggested as responsible for SC induction presented in Table II. This resulted in lower estimates of solid SC risk but unfortunately also in some deterioration of spinal target coverage. Such an optimization approach should thus preferably be accompanied by assessment of tumor control probability.

The risks of all long-term complications, cancer and non-cancer alike, were shown to be considerably lower for IMPT than for the photon techniques. Pediatric MB thus seems to be an obvious candidate to recommend for proton therapy if available. The risk estimation for the IMPT plans is made difficult by uncertainty in the RBE of SN produced by the proton beam. The conclusions for IMPT as favorable

<table>
<thead>
<tr>
<th>Parameter</th>
<th>3D CRT</th>
<th>RapidArc</th>
<th>IMPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>D&lt;sub&gt;min&lt;/sub&gt;</td>
<td>13.5 Gy</td>
<td>15.5 Gy</td>
<td>13.7 Gy</td>
</tr>
<tr>
<td>D&lt;sub&gt;max&lt;/sub&gt;</td>
<td>29.5 Gy</td>
<td>27.8 Gy</td>
<td>26.5 Gy</td>
</tr>
<tr>
<td>V&lt;sub&gt;95&lt;/sub&gt;</td>
<td>96.6%</td>
<td>95.2%</td>
<td>97.1%</td>
</tr>
<tr>
<td>V&lt;sub&gt;107&lt;/sub&gt;</td>
<td>4.5%</td>
<td>6.3%</td>
<td>0.2%</td>
</tr>
<tr>
<td>CI</td>
<td>0.61</td>
<td>0.73</td>
<td>0.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PTV craniospinal</th>
<th>PTV boost</th>
<th>CTV - Brain</th>
<th>CTV - Spinal</th>
</tr>
</thead>
<tbody>
<tr>
<td>D&lt;sub&gt;min&lt;/sub&gt;</td>
<td>13.5 Gy</td>
<td>45.1 Gy</td>
<td>21.5 Gy</td>
<td>21.3 Gy</td>
</tr>
<tr>
<td>D&lt;sub&gt;max&lt;/sub&gt;</td>
<td>29.5 Gy</td>
<td>57.3 Gy</td>
<td>25.0 Gy</td>
<td>28.6 Gy</td>
</tr>
<tr>
<td>V&lt;sub&gt;95&lt;/sub&gt;</td>
<td>96.6%</td>
<td>99.3%</td>
<td>99.3%</td>
<td>98.7%</td>
</tr>
<tr>
<td>V&lt;sub&gt;107&lt;/sub&gt;</td>
<td>4.5%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>17.1%</td>
</tr>
<tr>
<td>CI</td>
<td>0.61</td>
<td>0.61</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>D&lt;sub&gt;min&lt;/sub&gt;</td>
<td>15.5 Gy</td>
<td>41.1 Gy</td>
<td>19.6 Gy</td>
<td>21.0 Gy</td>
</tr>
<tr>
<td>D&lt;sub&gt;max&lt;/sub&gt;</td>
<td>27.8 Gy</td>
<td>59.4 Gy</td>
<td>27.7 Gy</td>
<td>25.9 Gy</td>
</tr>
<tr>
<td>V&lt;sub&gt;95&lt;/sub&gt;</td>
<td>95.2%</td>
<td>97.8%</td>
<td>97.3%</td>
<td>98.7%</td>
</tr>
<tr>
<td>V&lt;sub&gt;107&lt;/sub&gt;</td>
<td>6.3%</td>
<td>3.9%</td>
<td>7.5%</td>
<td>0.7%</td>
</tr>
<tr>
<td>CI</td>
<td>0.73</td>
<td>0.77</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>D&lt;sub&gt;min&lt;/sub&gt;</td>
<td>13.7 Gy</td>
<td>39.2 Gy</td>
<td>19.4 Gy</td>
<td>22.1 Gy</td>
</tr>
<tr>
<td>D&lt;sub&gt;max&lt;/sub&gt;</td>
<td>26.5 Gy</td>
<td>59.0 Gy</td>
<td>26.3 Gy</td>
<td>24.8 Gy</td>
</tr>
<tr>
<td>V&lt;sub&gt;95&lt;/sub&gt;</td>
<td>97.1%</td>
<td>98.6%</td>
<td>99.7%</td>
<td>99.5%</td>
</tr>
<tr>
<td>V&lt;sub&gt;107&lt;/sub&gt;</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.2%</td>
<td>0.0%</td>
</tr>
<tr>
<td>CI</td>
<td>0.89</td>
<td>0.90</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

CI: Conformity Index. Closer to unity equals better conformity, calculation details described elsewhere [45].
Radiobiological risk estimation in the treatment of pediatric medulloblastoma

over photon therapy were however shown in this study to hold even for high values of neutron RBE. This suggests that SN irradiation does not warrant the choice of photon therapy over proton therapy for MB. At least if proton therapy is given with IMPT with neutron contamination intensity as that used in the presented work, although the estimates are likely somewhat underestimated for the fields with range shifters. Passively scattered proton beams tend to yield more SN and needs to be investigated separately.

This study provided estimates and comparisons of solid secondary cancer risk and risks of noncancer late complications for three different radiotherapy techniques for treatment of pediatric MB. The uncertainties involved in the estimates should be considered, especially when interpreting the absolute risk estimates. The results of this study have shown that treating pediatric MB patients with 3D CRT would subject them to higher risks of several severe late complications compared to RA treatment. This shows the potential of intensity-modulated therapy in limiting the dose to important risk organs for these patients. The risk of SC induction was however estimated as higher if treated with RA compared to 3D CRT. This study also provided SC risk estimates as a function of attained age, highlighting the importance of considering at what ages different risks are most likely to occur.

Acknowledgements

This work was supported by funding from the Danish Childhood Cancer Foundation. The authors would like to thank chief physicians Catherine Rechnitzer and Karsten Nysom at the Department of Paediatric Oncology at Rigshospitalet, Copenhagen, for their support of this project. The authors would also like to thank everyone in the research group at the Radiation Oncology Department at Rigshospitalet, Copenhagen, for their valued input and discussions.

Declaration of interest: Per Munck af Rosenschöld has a research agreement with Varian Medical Systems. The authors alone are responsible for the content and writing of the paper.

References


[29] ICRP. Relative biological effectiveness (RBE), quality factor (Q), and radiation weighting factor (w(R)). Ann. ICRP: (Oxford: International Commission on Radiological Protection); 2003. Report No.: 92.


STUDY II
Life Years Lost—Comparing Potentially Fatal Late Complications After Radiotherapy for Pediatric Medulloblastoma on a Common Scale

N. Patrik Brodin, MSc1,2; Ivan R. Vogelius, PhD1; Maja V. Maraldo, MD3; Per Munck af Rosenschöld, PhD1,2; Marianne C. Aznar, PhD1,2; Anne Kil-Berthelsen, MD3; Per Nilsson, PhD4; Thomas Björk-Eriksson, MD, PhD4; Lena Specht, MD, DMSc3,5; and Søren M. Bentzen, PhD, DSc1,6

BACKGROUND: The authors developed a framework for estimating and comparing the risks of various long-term complications on a common scale and applied it to 3 different techniques for craniospinal irradiation in patients with pediatric medulloblastoma.

METHODS: Radiation dose-response parameters related to excess hazard ratios for secondary breast, lung, stomach, and thyroid cancer; heart failure, and myocardial infarction were derived from large published clinical series. Combined with age-specific and sex-specific hazards in the US general population, the dose-response analysis yielded excess hazards of complications for a cancer survivor as a function of attained age. After adjusting for competing risks of death, life years lost (LYL) were estimated based on excess hazard and prognosis of a complication for 3-dimensional conformal radiotherapy (3D CRT), volumetric modulated arc therapy (VMAT), and intensity-modulated proton therapy (IMPT).

RESULTS: Lung cancer contributed most to the estimated LYL, followed by myocardial infarction, and stomach cancer. The estimates of breast or thyroid cancer incidence were higher than those for lung and stomach cancer incidence, but LYL were lower because of the relatively good prognosis. Estimated LYL ranged between 1.90 years for 3D CRT to 0.28 years for IMPT. In a paired comparison, IMPT was associated with significantly fewer LYL than both photon techniques.

CONCLUSIONS: Estimating the risk of late complications is associated with considerable uncertainty, but including prognosis and attained age at an event to obtain the more informative LYL estimate added relatively little to this uncertainty.

Cancer 2012;000:000–000. © 2012 American Cancer Society.

KEYWORDS: life years lost, radiotherapy, late effects, secondary cancers, risk modeling.

INTRODUCTION
Cancer survivors are subject to elevated risks of adverse effects, such as cardiac disease, blindness, pneumonitis, neurocognitive impairment, and secondary cancers (SCs).1-3 Information from cohort studies can be used to model the risk of late complications, for example, to compare different treatment modalities or strategies.4,5 Risk estimates usually are expressed as a percentage excess risk or a relative risk (RR) compared with the general population. One problem is that different treatment options may give rise to different complications, and it is not straightforward to trade-off, say, an increased lifetime excess risk of severe late cardiac events with a decreased risk of inducing an SC.

In this report, we propose life years lost (LYL) as a tool for comparing multiple risks of potentially fatal late complications. This takes into account the age-dependent risk of a given late event as well as its prognosis. Furthermore, a treatment-related fatality occurring at a young age will cause a greater average loss of life expectancy than the same event occurring late in life. The measure is easy to interpret and can be used to prioritize between risks, for example, of cardiac events and SCs. In this study, LYL are applied to compare 3 craniospinal irradiation (CSI) techniques for 10 pediatric patients with medulloblastoma (MB).

MATERIALS AND METHODS

Concept Methodology
LYL are estimated from the age-specific excess hazard ratio ($h_{\text{excess}}$) of cancer survivors compared with the general population. To facilitate treatment-specific risk assessment, the dependency of $h_{\text{excess}}$ on the radiation dose, $D$, must be obtained for each of the studied endpoints. Also, the optimal dosimetric descriptor (eg, mean dose, median dose, or maximum...
After a heart failure diagnosis, however, the same patient
exposure that deriving
excess per year of attained age for
developing the studied complication.

To account for competing risks of death from the pri-
mary disease, treatment-related mortality, or noncancer-
related events, the probability of reaching age \( a \), \( S(a, s) \) was
estimated using survival curves from the Childhood Cancer
Survivor Study (CCSS) cohort relative to an age-matched
and sex-matched US general population.\(^7\) These data cover
30 years since diagnosis but were linearly extrapolated
beyond this period by assuming the same trend in survival
ratio between the CCSS cohort and the general population
(see Figure 1). The childhood cancer survival curve is nor-
malized to the survival from the primary disease assuming a
5-year survival rate of 80%. The sensitivity of LYL to this
extrapolation was tested by recalculating the survival curve,
assuming that the survival ratio between the CCSS cohort
and the general population remained constant after the last
empirical observation.

The age-dependent and sex-dependent life expectancy
(LE) after a complication at age \( a \) is estimated from empiri-
cal survival data assuming the same prognosis as that for a
spontaneous event extracted from national registries. Integrat-
ing the survival probability yields the LE after the corre-
spending event. The LYL attributable to a specific endpoint
occurring at age \( a \) is then the difference in LE relative to an
individual of the same age in the general population, condi-
tional on having survived until that age. Hence:

\[
LYL(D, e, s, a) = S_{\text{survivor}}(a, s) \cdot \dot{h}_{\text{excess}}(D, e, s, a) \\
\cdot (LE_{\text{gen.pop.}}(a, s) - LE_{\text{endpoint}}(a, s))
\]  

Figure 2 provides a flowchart illustrating the meth-
ology with a corresponding example of calculating the
LYL for a specific case. The total LYL attributable to each
endpoint can then be derived by integrating the LYL for
all attained ages after the age at exposure.

\[
LYL(D, e, s) = \int_{e+1}^{\infty} S_{\text{survivor}}(a, s) \cdot \dot{h}_{\text{excess}}(D, e, s, a) \\
\cdot (LE_{\text{gen.pop.}}(a, s) - LE_{\text{endpoint}}(a, s)) \, da
\]

The upper limit of the integral was taken as 110
years, because \( S_{\text{survivor}} \) for older ages was effectively zero.

**Application to Pediatric Medulloblastoma Patients**

Different radiotherapy treatment techniques were com-
pared with respect to LYL for 10 pediatric patients with
MB ages 4 to 15 years who received treatment during
2007 to 2009 with postsurgical chemotherapy and CSI
at our institution. Table 1 summarizes literature data on
radiation dose-response and effect of age at exposure,
attained age, and patient sex. Data from pediatric stud-
ies were used where possible; alternatively, data from
Hodgkin lymphoma survivors were used. The \( hr_{\text{excess}} \) as
a function of radiation dose is normalized to that of
unirradiated individuals, because the case-control design
does not allow for absolute risk estimates. This assumes
that deriving \( hr_{\text{excess}} \) from nonirradiated cancer patients
is identical to deriving it from the US general population.

Age-specific and sex-specific hazard rates and survival
data for thyroid, breast, and lung cancer were obtained from the SEER registry.\(^6\) Data for cardiac events
were obtained as hospital discharge rates from the Centers
for Disease Control and Prevention (CDC) database.\(^16\) After a heart failure diagnosis, however, the same patient
may be hospitalized repeatedly. Therefore, age-specific
mortality rates, \( \dot{h}_{\text{mort}} \), were obtained from the CDC data-
base and LYL estimated as:
Figure 2. This flow chart illustrates the life years lost (LYL) estimation methodology and a corresponding example of calculating the LYL for a specific patient. 3D CRT indicates 3-dimensional conformal radiotherapy; Gy, gray; max, maximum; DVH, dose-volume histogram; hr, hazard ratio; D, radiation dose; M:F, male:female ratio.

\[
LYL_{hearts} = S_{surv}(a, s) \cdot hr_{mortality}(D, e, s, a) \cdot LE_{gen.pop}(a, s) \tag{4}
\]

LYL_{hearts} = S_{surv}(a, s) \cdot hr_{mortality}(D, e, s, a) \cdot LE_{gen.pop}(a, s)

Treatment plans for the 10 patients with MB delivering 36 gray (Gy) to the craniospinal axis and 54 Gy to the posterior fossa were generated in accordance with published guidelines using 3-dimensional conformal therapy (3D CRT); volumetric modulated arc therapy (VMAT); and spot-scanned, intensity-modulated proton therapy (IMPT) with the Eclipse treatment planning system (version 8.9; Varian Medical Systems, Palo Alto, Calif). The 3D CRT plans consisted of 2 lateral opposed cranial fields to cover the whole brain and a spinal posterior-anterior field to cover the spinal canal extending caudally to the junction of spinal segments 2 and 3. The lateral extent of the 3D CRT spinal field was the planning target volume plus 10 mm, because this best represented the...
set-up from the treatment records of the patients. The boost plans were created using 4 fields to cover the posterior fossa. The VMAT plans were generated using the RapidArc (Varian Medical Systems) implementation with a 360-degree cranial arc and one or two 140-degree spinal arcs. The spinal arcs were not planned as full rotations to avoid irradiating through the arms and ventral parts of the patient. The VMAT boost plans consisted of a 220-degree arc covering the posterior fossa without incident irradiation through the anterior part of the head.

The IMPT plans were generated using 3 fields incident from the posterior direction. One field was set to cover the caudal part of the spinal canal, and the other 2 fields were set to cover the brain and remaining part of the spinal canal. The IMPT boost plans consisted of a single posterior field covering the posterior fossa. The clinical target volume-to-planning target volume margins for the VMAT and IMPT plans were 5 mm for the cranial part and 7 mm for the spinal part. Note that changing the margins would change the doses to the organs at risk (OARs) close to the target and, thus, would affect the relative comparison between treatment techniques.

Further details regarding treatment planning and margins have been described elsewhere. It should be noted that, although the dosimetric input in the current study was based on Brodin et al, the risk estimates provided here were based on a different methodology and dose-response data. The same treatment margins were applied for IMPT as for the VMAT technique, although, in practice, larger margins are sometimes applied with proton therapy because of the range uncertainty in proton dose deposition. OARs were delineated by an experienced radiologist (A.K.-B.).

Estimated, organ-specific, secondary neutron doses were obtained from a spot-scanned proton beam that was simulated using Monte Carlo methods on a pediatric mathematical phantom receiving craniospinal proton irradiation. The induction of late effects attributable to secondary neutrons is subject to large uncertainty. Therefore, as a conservative approach, the recommended relative biologic effectiveness of neutrons compared with photons was multiplied by a factor of 5 in our estimates. Even so, the secondary neutrons inherent to proton irradiation will contribute to added uncertainty in our estimates. Also, because our estimates are based on a simulated, spot-scanned proton beam, they may not be valid for a passively scattered beam.

Dose-response relations for the heart, breast, lung, and stomach endpoints were approximately linear, so the mean dose was used for these calculations. The dose-
response for secondary thyroid cancer, however, is best modeled as a bell-shaped function. Consequently, the differential dose-volume histogram of the thyroid was weighted to obtain an organ equivalent dose.

\[ h_{\text{excess, thyroid}} = \frac{1}{V} \sum_{i=1}^{N} v_i \beta_1 D_i \cdot \exp(-\beta_4 D_i^2) \]  

(5)

where \( v_i \) is the fractional volume of the thyroid receiving dose \( D_i \), and \( N \) is the number of voxels in the total volume \( V \). The \( \beta_1 \) parameter describes the linear ascending part of the dose-response, and \( \beta_4 \) is the exponential-quadratic descending part (see Table 1).

**Statistical Analysis**

Uncertainty in estimating LYL comes mainly from the dose-response parameters. To obtain a robust comparison of LYL between treatment modalities, a paired difference Monte Carlo method was used. Samples were randomly drawn from a log-normal distribution with mean and 95% confidence interval (CI) corresponding to the published dose-response data. Thus, we sampled over the uncertainty in dose-response for each endpoint derived from the corresponding epidemiological studies. However, this does not include systematic components; for example, if the true functional form of the dose-risk curve is different from the assumed form or if the dose descriptor used here is not the best fit to the outcome data, then such uncertainties cannot be represented by the uncertainty in the model parameters.

For each endpoint, the difference in LYL between pairs of modalities was calculated, and the overall LYL with 95% CI were extracted by inverse variance weighting. To avoid underestimation of the variance because of the small number of patients in this study, a bootstrapping procedure was applied in which 10,000,000 samples from the 10 patients were drawn with replacement. The average point estimate and CI was calculated as stated above for each sample, and a normal distribution was matched to the result. Finally, a sample was randomly drawn from each of these 10,000,000 normal distributions. The final difference in LYL with 95% CI was taken as the mean and percentiles 2.5 to 97.5 of the randomly drawn samples. This provided estimates of the uncertainty in the paired comparison between modalities, resulting from uncertainty in the parameters derived from the published dose-response data.

**RESULTS**

Age-specific hazard rates for cancer and cardiac events in the general population were not published for individuals aged \( \geq 85 \) years. It was assumed that these rates are constant for all those aged \( >90 \) years and equal to the rates published for the group aged \( \geq 85 \) years. This assumption has a negligible effect on the estimated LYL.

Figure 3 illustrates the estimated time of onset between endpoints, in which the peak in attributable LYL appears earlier for breast cancer than for lung cancer. The predominant contributor to the LYL for the VMAT plans is lung cancer followed by stomach cancer and myocardial infarction (see Figure 4). The estimated early onset and relatively high incidence of secondary thyroid cancer (data not shown) is offset by its favorable prognosis when calculate the LYL. Figure 4 provides total LYL estimates according to Equation 3 for the different treatment techniques.

The treatment-related mean (95% CI) LYL differences between VMAT and IMPT was 1.09 years (95% CI, 0.80-1.42 years) and 0.37 years (95% CI, 0.23-0.52 years) between 3D CRT and VMAT. LYL attributable to SC was higher for VMAT than for 3D CRT, reflecting the spread of radiation dose to OARs typical of intensity-modulated treatment. Conversely, the LYL attributable to cardiac events were lower for VMAT than for 3D CRT, reflecting the considerably reduced mean heart dose (7.3 Gy vs 18.9 Gy). The difference between 3D CRT and VMAT depends on the relative weight of cardiac events compared with SC events and, thus, is subject to a
systematic uncertainty that is not included in the statistical uncertainty and should be cautiously interpreted. The width of the spinal treatment field in 3D CRT for MB affects the risk of radiation-induced heart failure, because it directly affects the mean heart dose.4 The same holds true for the mean lung dose. Hence, the relative merits of these techniques depend critically on the margins used.

The sensitivity analysis of extrapolating the survival curve of the CCSS cohort to older ages indicated that assuming a constant rather than a linearly decreasing ratio after the last follow-up yielded 12% higher LYL estimates for all 3 treatment techniques but did not affect their relative merits. The paired samples statistical test revealed a small but significant difference in LYL between 3D CRT and VMAT, although it should be noted that any systematic uncertainty that may change the relative weights of SCs compared with cardiac events is not considered in the statistical analysis, and there is a risk of a systematic error stemming from a mismatch between the assumed shape and the true shape of the underlying dose-response relation. The strength of the paired Monte Carlo test is that only differences in risks are considered, rendering the conclusion insensitive to uncertainties in parameters that describe a monotonous dose-response curve. However, as most statistical tests, the method relies on the assumption that the dose descriptor (typically, the mean dose) is indeed the correct descriptor and that the dose response has the correct parameterization. Figure 5 illustrates the paired Monte Carlo test with bootstrapping for this comparison. The LYL because of secondary thyroid cancer were less for 3D CRT compared with VMAT despite a higher thyroid dose because of the bell-shaped dose-response relation.

Figure 6 illustrates the lifetime cumulative risk of developing an SC and the corresponding LYL, clearly illustrating the impact of disease prognosis on LYL. The most frequent SCs were breast cancer and thyroid cancer; however, the LYL attributable to these malignancies were small compared with the LYL attributable to lung cancer. The mean lifetime cumulative risk of developing any of the studied SCs was 33%, 40%, and 18% for 3D CRT, VMAT, and IMPT, respectively.

**DISCUSSION**

A framework was developed for estimating the LYL attributable to late complications of radiation therapy, such as SC and cardiac events. The advantage of the LYL estimate over assessment of lifetime cumulative risks is that the time to event and the prognosis are taken into account. The uncertainty in relating radiation dose to excess hazard dominates the uncertainty of both the LYL estimate and estimates of lifetime cumulative risk. The additional data needed for estimating LYL are extracted from population-based registries and, thus, are known with comparatively high precision. A required assumption is that the prognosis after SCs or radiation-induced non-malignant toxicity is similar to that observed after an event of nonradiation-related etiology. Consequently, LYL are easier to interpret when assessing the relative merits of alternative radiation therapy plans than lifetime cumulative risks and is associated with only a minor increase in uncertainty. Although it is informative, the LYL measure is a result of modeling various risks based on epidemiological studies and, thus, should be used with caution at the individual patient level. Further validation of the model is required, and the conclusions regarding the relative merits of radiation modalities should be tested in independent data sets.

In the current study, dose-response relations were extracted from large clinical series and from hazard rates documented for the US general population. When interpreting case-control study data for an endpoint, we assumed that $h_{excess}$ for an unirradiated cancer survivor was identical to that of an age-matched and sex-matched individual in the general population, thus assuming no impact of chemotherapy or genetic predisposition. A similar problem was identified by Travis et al20 when estimating cumulative absolute risks of secondary breast cancer after treatment for Hodgkin lymphoma based on a nested case-control study.8 Those authors proposed treating the
RR from the case-control study as an internal risk, estimating an external risk for the same cohort related to the general population breast cancer incidence, then combining these to estimate absolute risks. This approach requires individual patient-level data from the case-control study. In addition, it is unclear whether age at exposure and attained age significantly affect the late complication risk. The estimates provided in Figure 4 primarily represent the possibility of affecting the LYL by redistribution of radiation dose. Including nonradiation-related risks more closely represents the absolute treatment-induced LYL of a cancer survivor. It is indeed possible to include risks like anthracycline-related cardiac complications in the presented framework, but the required data could not be extracted from the current literature.

The statistical uncertainty illustrates the need for higher quality dose-dependent risk data reported as standardized incidence ratios (SIRs) relative to the general population. However, the dose-dependence of complication risks and the risk-modifying effect of patient-related factors often are not reported in sufficient detail for optimal estimation of LYL. In the current analysis, the dependence on age at exposure was only available for the thyroid cancer estimates. Ideally, fitting parameters with CIs and the functional form of the multivariate model should be reported (for example, see the report by Bhatti et al13).

The translation of risk estimates derived from large epidemiological studies with retrospective dosimetry into estimates for the patients in our study contributes to the overall uncertainty. This is especially true for endpoints in which the dose response was based on adult data, because

---

**Figure 5.** These are results from statistical analyses of the difference in life years lost (LYL) with 3-dimensional conformal radiotherapy (3D CRT) and volumetric modulated arc therapy (VMAT). The mean differences with corresponding 95% confidence intervals (CIs) are listed for each patient from the Monte Carlo sampling along with the resulting distribution of difference estimates obtained from bootstrapping and final estimates of mean and 95% CI.
a general concern is that a developing child may be more susceptible to radiation-induced cancers. Also, long-term follow-up studies like the CCSS reflect the risk of treatment given some 20 to 30 years ago, making it difficult to extrapolate the results to modern treatment techniques like VMAT or IMPT. These uncertainties mean that validation against independent data sets is crucial if an LYL measure is to be trusted for clinical decision support.

A recent publication\textsuperscript{21} estimated the nonrelapse-associated LYL in MB survivors at 4.3 years based on published mortality data from the CCSS.\textsuperscript{7} This value is larger than our estimates, probably because only a few causes of mortality are considered here. The estimated lifetime cumulative risks of SC of 33\% for 3D CRT and 40\% for VMAT are roughly in agreement with other estimates for pediatric patients with MB at 20\% to 55\%.\textsuperscript{22,23} It is noteworthy that Mertens et al\textsuperscript{7} reported no cardiac deaths among MB survivors, possibly because the patients they studied had not reached an age at which the risk of cardiac events in the general population becomes apparent.

The current study assumes a similar prognosis after a treatment-induced event and a spontaneous event. Cancer survivors may be followed more closely than the general population, possibly leading to earlier detection of an SC. The SC also may be of a different phenotype, more or less difficult to manage, than its spontaneous counterpart. Treatment-related cardiac complications may have a worse prognosis because of the relatively high prevalence of cardiac risk factors among cancer survivors.\textsuperscript{24} Such nuances can easily be incorporated into the LYL estimation if the appropriate data become available.

A logical, but challenging, extension of the LYL concept would be the inclusion of late relapse of the primary disease. Curing the primary disease must be the highest priority in cancer care, and failure to do this would yield a large number of LYL. Including estimates of tumor control requires reliable clinical dose-response data and, ideally, a quantification of the loss of tumor control if part of the target volume is under treated. Another extension of our model would be to include nonlethal late complications that affect quality of life to estimate quality-adjusted LYL (QALYL). QALYL theoretically would allow optimization of both health-related quality of life and life expectancy and thereby allow assessment of the cost-benefit of advanced therapy options, such as proton therapy.\textsuperscript{25}

Finally, LYL or QALYL could form the basis for radiotherapy plan optimization.

In summary, LYL estimates attributable to late effects are objective, easy to interpret, and take prognosis and time to the event into account when comparing alternative treatment options. Our current limited knowledge of treatment-induced late effects does limit the accuracy of long-term risk estimates. However, several large clinical studies are in progress that will reduce the uncertainty of dose-response and clinical risk factor data and thereby improve the accuracy of LYL estimates.

**FUNDING SOURCES**

This work was supported by a grant from the Danish Child Cancer Foundation to N. Patrik Brodin. Soren M. Bentzen acknowledges support from the National Cancer Institute (grant 2P30 CA 014520-34).

**CONFLICT OF INTEREST DISCLOSURES**

Per Munck af Rosenschöld has a research agreement with Varian Medical Systems.

**REFERENCES**


STUDY III
Estimated clinical benefit of protecting neurogenesis in the developing brain during radiation therapy for pediatric medulloblastoma

Malin Blomstrand, N. Patrik Brodin, Per Munck af Rosenschöld, Ivan R. Vogelius, Gaspar Sánchez Merino, Anne Kiil-Berthlesen, Klas Blomgren, Birgitta Lannering, Søren M. Bentzen, and Thomas Björk-Eriksson

Department of Oncology, Sahlgrenska University Hospital, Gothenburg, Sweden (M.B.); Centre for Brain Repair and Rehabilitation, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden (M.B.); Radiation Medicine Research Center, Department of Radiation Oncology, Rigshospitalet, Copenhagen, Denmark (N.P.B., P.M.R., I.R.V., A.K.B., S.M.B.); Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark (N.P.B., P.M.R.); Department of Medical Physics and Radiation Protection, Txagorritxu Hospital, Vitoria, Spain (G.S.M.); Department of Clinical Physiology and Nuclear Medicine, Centre of Diagnostic Investigations, Rigshospitalet, Copenhagen, Denmark (A.K.B.); Karolinska Institutet, Department of Women’s and Children’s Health, Karolinska University Hospital Q2:07, Stockholm, Sweden (K.B.); Department of Pediatric Oncology, The Queen Silvia Children’s Hospital, University of Gothenburg, Gothenburg, Sweden (K.B.); Center for Brain Repair and Rehabilitation, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden (K.B.); Department of Pediatric Oncology, The Queen Silvia Children’s Hospital, University of Gothenburg, Gothenburg, Sweden (B.L.); Department of Human Oncology, University of Wisconsin Medical School, Madison, Wisconsin (S.M.B.); Department of Oncology, Skåne University Hospital and Lund University, Lund, Sweden (T.B.E.)

We sought to assess the feasibility and estimate the benefit of sparing the neurogenic niches when irradiating the brain of pediatric patients with medulloblastoma (MB) based on clinical outcome data. Pediatric MB survivors experience a high risk of neurocognitive adverse effects, often attributed to the whole-brain irradiation that is part of standard management. Neurogenesis is very sensitive to radiation, and limiting the radiation dose to the hippocampus and the subventricular zone (SVZ) may preserve neurocognitive function. Radiotherapy plans were created using 4 techniques: standard opposing fields, intensity-modulated radiotherapy (IMRT), intensity-modulated arc therapy (IMAT), and intensity-modulated proton therapy (IMPT). Mean dose to the hippocampus and SVZ (mean for both sites) could be limited to 88.3% (range, 83.6%–91.0%), 77.1% (range, 71.5%–81.3%), and 42.3% (range, 26.6%–51.2%) with IMAT, IMRT, and IMPT, respectively, while maintaining at least 95% of the prescribed dose in 95% of the whole-brain target volume. Estimated risks for developing memory impairment after a prescribed dose of 23.4 Gy were 47% (95% confidence interval [CI], 21%–69%), 44% (95% CI, 21%–65%), 41% (95% CI, 22%–60%), and 33% (95% CI, 23%–44%) with IMAT, IMRT, and IMPT, respectively, while maintaining at least 95% of the prescribed dose in 95% of the whole-brain target volume. Estimated risks for developing memory impairment after a prescribed dose of 23.4 Gy were 47% (95% confidence interval [CI], 21%–69%), 44% (95% CI, 21%–65%), 41% (95% CI, 22%–60%), and 33% (95% CI, 23%–44%) with opposing fields, IMAT, IMRT, and IMPT, respectively. Neurogenic niche sparing during cranial irradiation of pediatric patients with MB is feasible and is estimated to lower the risks of long-term neurocognitive sequelae. Greatest sparing is achieved with intensity-modulated proton therapy, thus making this an attractive option to be tested in a prospective clinical trial.

Corresponding Author: Thomas Björk-Eriksson, Department of Oncology, Skåne University Hospital and Lund University, SE-221 85 Lund, Sweden (thomas.bjork-eriksson@skane.se).
Keywords: CNS, medulloblastoma, neurocognitive sparing, radiotherapy, risk modeling.

Radiotherapy is one of the most effective therapeutic modalities for malignant central nervous system (CNS) tumors. Medulloblastoma (MB) accounts for about 20% of CNS tumors in children, and the peak incidence is at 5 years of age. The prognosis of MB, which is a primitive neuro-ectodermal tumor (PNET) located in the cerebellum or the fourth ventricle, improved considerably with the introduction of adjuvant radiotherapy. This success has generated a growing population of children surviving their cancer. Irradiation of the CNS is, however, associated with a risk of severe adverse effects, including neurocognitive sequelae. Younger age at treatment is correlated with more severe cognitive deficits. The detailed pathogenesis of cognitive dysfunction after radiotherapy is yet unknown, but several mechanisms likely play a role. Post-irradiation MRI reveals multiple changes, including white matter microstructure disruptions, decreased size of corpus callosum and subregions, and abnormal hippocampal development. In mammals, neurogenesis occurs at 2 major sites, the subgranular zone (SGZ) in the dentate gyrus of the hippocampus and the subventricular zone (SVZ) of the lateral ventricles. The stem and progenitor cells in these niches are sensitive to irradiation, and recent discoveries in neural stem cell biology and brain plasticity have provided clues toward a deeper understanding of the effects of ionizing radiation of the developing brain.

Irradiation of the CNS is, however, associated with a risk of severe adverse effects, including neurocognitive sequelae. Younger age at treatment is correlated with more severe cognitive deficits. The detailed pathogenesis of cognitive dysfunction after radiotherapy is yet unknown, but several mechanisms likely play a role. Post-irradiation MRI reveals multiple changes, including white matter microstructure disruptions, decreased size of corpus callosum and subregions, and abnormal hippocampal development. In rodents, neurogenesis has been shown to be important for hippocampal-dependent memory formation. Several human studies demonstrate a relationship between absorbed dose to the brain and cognitive outcome. and recent studies show more specifically a correlation between temporal lobe irradiation and neurocognitive sequelae. Thus, because neurogenesis is important for hippocampal-dependent memory and the hippocampus is situated in the temporal lobe, it seems reasonable to hypothesize that the hippocampus is the main critical structure for radiotherapy-related cognitive function impairment. The role of SVZ irradiation for cognitive outcome shown is less clear; however, because of the proposed regenerative features of neurogenesis, the SVZ is included as an organ at risk (OAR) in this study.

New radiotherapy techniques, such as intensity-modulated radiotherapy (IMRT), intensity-modulated arc therapy (IMAT), and intensity-modulated proton therapy (IMPT), have facilitated the delivery of highly conformal dose distributions. Defining the hippocampus and the SVZ as OARs on a pre-radiotherapy magnetic resonance (MR) examination fused with CT scan facilitates reducing the dose to these regions during craniospinal irradiation (CSI). This would be particularly relevant for the hippocampus, because this region is important for memory function. With a dose prescription of 23.4–36 Gy, which is used for patients with MB, neurocognitive dysfunction is reported to be a common adverse effect. If a significant reduction in dose to the OARs can be achieved, a reduction in late cognitive adverse effects would be expected. A steep dose gradient would be needed to achieve a homogeneous dose to the rest of the brain while sparing the hippocampus and SVZ.

Inverse-planned intensity-modulated therapy aims at optimizing the dose distribution inside the patient’s body, guided by dose-volume objectives for tumor and OARs. The dose distribution is thus shaped around the target volume, often with a steep dose gradient to the neighboring tissue. The trade-off between treating the target to a sufficient and homogeneous dose and avoiding the OARs can be manipulated by the choice of planning dose-volume objectives. Conventional therapy, IMRT, and IMAT use photon beams for radiation dose deposition. Protons deposit their energy in tissue in a very different fashion than do photons, and their main characteristic is the sharp dose gradient at the distal edge of the beam. The IMPT technique therefore allows intensity modulation with a sharper dose fall-off, compared with the photon techniques.

Our aim with this retrospective dose planning study, focusing on the cranial component of the CSI course, was to evaluate how much modern radiation therapy techniques can reduce the absorbed dose to the hippocampus and SVZ and still treat the rest of the brain to an adequate radiation dose. We also evaluated the potential clinical benefit of this dose reduction based on dose-response data from a large clinical series with long-term follow-up on neurocognitive function of pediatric patients treated with radiation. In this study, we intend to explore the technological foundation for a prospective clinical trial based on dose-sparing of the hippocampus and SVZ.

Materials and Methods

Patients and Treatment Planning

Six patients with MB who all received CSI during 2002–2007 at Sahlgrenska University Hospital, Gothenburg, Sweden, were re-planned. Their age at time of treatment ranged from 6 to 11 years (median, 7.5 years). The clinical target volume (CTV) in this study comprised the whole brain, disregarding the spinal part of the target. The dose contribution to the hippocampus and SVZ from the boost treatment was assumed to be negligible in the present analysis. In reality, this dose contribution could be important depending on the size and location of the primary tumor, the treatment strategy used (treating the whole posterior fossa or only the tumor bed with a margin), and the treatment technique. Thus, the estimates presented in this study will apply to cases in which the assumption of a zero dose contribution to the hippocampus and SVZ from the boost is reasonable.

The OARs consisted of the hippocampus, the SVZ, and the eyes, which were delineated in each of the patients based on fused T1 and T2 MRI and CT scans. Target and OARs were delineated by an experienced neuroradiologist. The SVZ was defined as the lateral wall of the lateral ventricles with a margin of 2 mm.
Figure 1 illustrates the SVZ and hippocampus overlaid on a transversal MRI scan. The prescribed dose was set to 36 Gy or 23.4 Gy, in 1.8 Gy/fraction. All treatment plans were generated using the Eclipse treatment planning system, version 10 (Varian Medical Systems). Treatment planning was performed with the aim of minimizing the mean absorbed dose to the hippocampus and the SVZ without compromising CTV coverage. Four different radiotherapy delivery techniques were tested in this study, as shown in Fig. 2: 2 opposing cranial fields (which is still commonly used for cranial irradiation during the CSI course), IMRT with 7 fields, IMAT with 3 arcs (2 360 degree arcs and 1 noncoplanar 180 degree arc), and spot-scanned IMPT with 3 incident fields.

For the opposing field technique, the OAR sparing was limited to partial shielding of the eyes and the oral cavity. To ensure that the results of the treatment planning in this study were as user-independent as possible, we defined a fixed set of dose-volume objectives for the 3 inversely-optimized techniques: IMRT, IMAT, and IMPT. The objectives were defined in relation to 4 different levels of OAR sparing priority as shown in Table 1, with the intent of finding how the CTV radiation dose homogeneity was affected by the different levels of OAR sparing. We derived a linear correlation between the mean dose received by the hippocampus and the SVZ, further referred to as neurocognitive OAR dose, and volume of the CTV receiving at least 95% of the prescribed dose, the V95. We then estimated what OAR sparing could be achieved, for each individual patient, with the different techniques if the V95 was set to be at least either 98% or 95%, with the mean target dose fixed at 100% of the prescribed dose for all techniques. By doing so, we attempted to obtain an objective measure of how much the different techniques were able to spare the neurocognitive OARs and how this was affected by the limit chosen as the acceptable target coverage.

**Estimating the Risk of Neurocognitive Impairment**

A dose-response relationship for neurocognitive outcome was, until recently, available from animal studies only. However, Armstrong et al. provided dose-response data based on long-term survivors of childhood CNS malignancies. The authors found a correlation between radiation dose to the temporal lobe, while controlling for dose to other parts of the brain, and the risk of reduced task efficiency, organization, and memory. The assumption in the present study is that sparing the hippocampus and SVZ would be as effective as sparing the whole temporal lobe in terms of reducing neurocognitive adverse effects. Stratifying their data into a separate MB/PNET group, Armstrong et al. published odds ratios (ORs) with 95% confidence intervals (CIs), per 10 Gy increase in temporal lobe dose, for developing various neurological sequelae. From these ORs and the baseline risk of impairment with no temporal lobe irradiation, we derived logistic dose-response functions as follows:

\[
\begin{align*}
\text{ORD} &= \frac{p_D}{1-p_D} = \frac{OR_{10}}{1+OR_{10}} \\
\Rightarrow p_D &= \frac{OR_{10}}{\left(\frac{1}{p_0} - 1\right) + OR_{10}}
\end{align*}
\]

where \(D\) is the dose in Gy, \(OR_{10}\) is the corresponding OR at 10 Gy, \(p_0\) is the baseline risk of impairment at zero dose, and \(p_D\) is the risk of impairment at dose \(D\). The baseline risk was estimated from patients from the whole cohort who had not received any temporal lobe irradiation, not only from the stratified group. Separate estimates were not given, possibly because there were...
only 5 patients in the MB/PNET group who did not receive any cranial irradiation, which was too few to obtain a reliable baseline estimate. Applying the neurocognitive OAR doses from the treatment planning to the derived dose-response relations, risks of neurocognitive impairment between radiotherapy techniques were estimated for 2 prescribed dose levels: 23.4 Gy and 36 Gy.

**Table 1.** Dose-volume objectives used in the inversely-optimized treatment planning. The same priority was applied for the eyes, hippocampus, and SVZ in each of the various OAR settings

<table>
<thead>
<tr>
<th>Structure</th>
<th>Volume (%)</th>
<th>Dose (Gy)</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body</td>
<td>0</td>
<td>37.5</td>
<td>250</td>
</tr>
<tr>
<td>CTV</td>
<td>0</td>
<td>36.5</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>35.5</td>
<td>225</td>
</tr>
<tr>
<td>OAR setting 0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>OAR setting 1</td>
<td>0</td>
<td>5</td>
<td>65</td>
</tr>
<tr>
<td>OAR setting 2</td>
<td>0</td>
<td>5</td>
<td>120</td>
</tr>
<tr>
<td>OAR setting 3</td>
<td>0</td>
<td>5</td>
<td>160</td>
</tr>
</tbody>
</table>

**Statistical Analysis**

The largest source of uncertainty in this study was the OR estimates from Armstrong et al. on which we based our dose-response functions. To test whether the risks of impairment between treatment techniques were significantly different, a paired random number (Monte Carlo) test comparing the OR between 2 treatment techniques was performed. Samples were drawn randomly from log-normal distributions corresponding to the mean and 95% CI of the dose-response parameters. For each of the different neurocognitive end points, the OR between techniques with 95% CI was calculated by inverse variance weighting. To account, to some extent, for the possible underestimation of the variance resulting from the small number of patients, a bootstrapping procedure was applied. Ten million samples of the 6 patients were drawn with replacement. A mean point estimate OR with CI was then calculated for each of the 10 million samples and a normal distribution matched to each one. Finally, 1 sample was randomly drawn from each of the 10 million distributions, giving the final OR and 95% CI as the mean and 2.5–97.5 percentile of the randomly drawn samples.

**Results**

There was a clear effect on CTV coverage when tightening the OAR dose constraint in the treatment planning process, as shown in Fig. 3. The CTV coverage was least affected for the IMPT plans, suggesting that the OARs can be spared to a greater extent with the proton technique. Of the highly conformal photon techniques, IMRT was slightly more effective than IMAT at sparing the neurocognitive OARs.

On the basis of the data in Fig. 3, it was deemed reasonable to describe the correlation between CTV V95 and neurocognitive OAR dose by a simple linear relation. The neurocognitive OAR doses, corresponding to CTV V95 equal to either 98% or 95%, were calculated using the linear regressions in Fig. 3 and are given in Table 2. These doses thus represented the lowest neurocognitive OAR dose achievable, with the different techniques, for a CTV coverage of V95 at either 98% or 95%.

**Table 2.** Mean doses (range) to neurocognitive organs at risk represented as percentage of the prescribed treatment dose.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Hippocampus (%)</th>
<th>SVZ (%)</th>
<th>Neurocognitive OAR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CTV V95 = 98%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protons</td>
<td>77.0 (71.0–80.6)</td>
<td>79.2 (71.6–87.0)</td>
<td>78.0 (71.3–81.1)</td>
</tr>
<tr>
<td>IMRT</td>
<td>89.7 (87.0–92.7)</td>
<td>90.5 (86.5–93.3)</td>
<td>90.0 (86.9–93.0)</td>
</tr>
<tr>
<td>IMAT</td>
<td>97.2 (95.9–98.4)</td>
<td>98.7 (96.6–101.7)</td>
<td>97.8 (96.2–99.8)</td>
</tr>
<tr>
<td>Opposing fields</td>
<td>98.0 (97.2–99.3)</td>
<td>100.4 (99.4–101.3)</td>
<td>99.2 (98.9–100.0)</td>
</tr>
<tr>
<td></td>
<td>CTV V95 = 95%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protons</td>
<td>41.9 (25.9–50.3)</td>
<td>42.7 (27.4–51.5)</td>
<td>42.3 (26.6–51.2)</td>
</tr>
<tr>
<td>IMRT</td>
<td>77.1 (72.5–81.1)</td>
<td>77.2 (70.6–81.5)</td>
<td>77.1 (71.5–81.3)</td>
</tr>
<tr>
<td>IMAT</td>
<td>87.1 (82.9–89.4)</td>
<td>89.6 (84.3–94.2)</td>
<td>88.3 (83.6–91.0)</td>
</tr>
<tr>
<td>Opposing fields</td>
<td>98.0 (97.2–99.3)</td>
<td>100.4 (99.4–101.3)</td>
<td>99.2 (98.9–100.0)</td>
</tr>
</tbody>
</table>

Note: The neurocognitive OAR dose was taken as the average dose of the hippocampus and the SVZ.
The logistic dose-response relationships shown in Fig. 4 were derived from the ORs and baseline risks in Armstrong et al. according to Equation 1. Figure 5 shows the estimated incidence of neurological sequelae based on the doses in Table 2 and the derived dose-response relationships.

As shown in Fig. 5, the risks of developing various neurological impairments were estimated to be lower with IMPT, compared with photon therapy. In addition, relaxing the CTV coverage constraint has a large impact on the possible neurocognitive function sparing with IMPT but little impact for the photon techniques.

The clinical relevance of low radiation doses to the CNS remains controversial, with a few studies reporting measurable decline in mental functioning after doses as low as 2 Gy and below. However, there is no doubt that the high doses of radiation prescribed for treating MB lead to neurocognitive deficits and are likely to influence later academic achievements and social life. Sparring the entire temporal lobe from irradiation would considerably reduce the dose to a large part of the target volume in MB and, thereby, likely increase the risk of relapse. Here, we assume that sparing the hippocampus and SVZ provides the same cognitive function sparing, but with the ability of maintaining a
CTV coverage of V95 at either 95% or 98%. We show that advanced radiation therapy techniques can reduce the dose to the hippocampus and SVZ without compromising the dose coverage of the whole-brain target volume.

To put our dosimetric findings into a clinical context, the risks of developing various neurocognitive sequelae were estimated, based on data from long-term follow-up of pediatric MB survivors. IMPT is superior to the conventional technique and also significantly better than IMRT and IMAT. When the CTV V95 constraint is relaxed from 98% to 95%, this gives a benefit mainly with the IMPT technique. In practice, however, this could partly be offset by the risk involved in lowering the target coverage constraints for IMPT because of the sharp dose fall-off at the distal field edge.

Slightly lower risk of impairment is estimated for IMRT than for IMAT but with wide CI, resulting from uncertainty in the dose-response parameters. However, the paired Monte Carlo test shows that ORs between techniques are all significantly different at the 95% level (Table 3). The high level of significance in the paired test reflects that, in each comparison, the estimates for all 6 patients favored the same technique, albeit to varying degrees. For a linear correlation between dose and the logarithm of the OR as assumed by Armstrong et al., a lower dose will always give a lower OR as long as the slope is positive (OR, >1). Consequently, the paired comparison of techniques circumvents the effect of uncertainty in the magnitude of a positive slope of the dose-response curve. Indeed, any positive dose-response function will retain the relative ranking of the plans. However, systematic errors could affect the comparison, for example, if mean dose is a poor predictor of toxicity or if there is a negative dose-response over a range of dose. Such systematic effects are not accounted for in the CI in Table 3. Furthermore, the small number of patients could lower the generalizability of our findings to a larger patient cohort.

Our estimated risks of neurological impairment are based on translating results from a study by Armstrong et al. based on long-term survivors of pediatric CNS malignancies. Our estimates are, thus, subject to the limitations of that study, such as the collection of data through questionnaires designed for neurocognitive function estimation. There are likely also some uncertainties in the radiation dosimetry, because this was based on retrospective evaluation of individual radiotherapy records and the fact that the majority of these patients were treated with older radiation delivery techniques. Despite these caveats, the data in this study are based on a large patient material followed up for a long time and with complete records of cranial radiotherapy for the 818 patients included. They also stratified patients specifically into a MB/PNET group, making the application in our study suitable.

The dose-response relationship between temporal lobe irradiation and neurocognitive impairment shown by Armstrong et al. was not limited to the MB/PNET group. For the whole patient cohort, the risk of task efficiency impairment was 24.0%, 34.7%, 48.3%, and 47.3% for temporal lobe doses of 0 Gy, 30–50 Gy, and >50 Gy, respectively. The corresponding risks of organizational impairment were 12.3%, 12.2%, 17.0%, and 22.6% and 24.6%, 33.3%, 45.1%, and 51.4% for impaired memory function. In health-related quality of life estimates they saw a correlation between temporal lobe irradiation and social functioning, physical limitations, and general health difficulties. Armstrong et al. also showed that patients in the MB/PNET group have a steeper temporal lobe dose-response, compared with survivors of other CNS tumors. ORs per 10 Gy–dose increase were 2.95 (95% CI, 1.66–5.22), 2.21 (95% CI, 1.04–4.70), and 1.45 (95% CI, 0.91–2.30) for task efficiency, organization, and memory, respectively, in the MB/PNET group and 1.10 (95% CI, 1.00–1.21), 1.12 (95% CI, 0.99–1.26), and 1.14 (95% CI, 1.03–1.25) for all other CNS tumors. The authors stipulate that a possible contributor to the substantial neurocognitive impairment in MB survivors could be the high-dose posterior fossa boost, the hypothesis being that this could cause loss of supratentorial connections between the cerebellum and the frontal region of the brain, which might affect executive functioning. This could possibly bias the baseline estimates of neurocognitive function at low temporal lobe doses but will probably not affect the steepness of the dose-response.

### Table 3. Estimates of odds ratios between the different treatment techniques and the 2 different dose prescriptions with corresponding 95% confidence intervals (CIs)

<table>
<thead>
<tr>
<th>Task efficiency</th>
<th>Organization</th>
<th>Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>36 Gy</td>
<td>23.4 Gy</td>
</tr>
<tr>
<td>Protons/IMRT V95 = 98%</td>
<td>0.64 (0.58–0.71)</td>
<td>0.75 (0.70–0.80)</td>
</tr>
<tr>
<td>Protons/IMRT V95 = 95%</td>
<td>0.28 (0.23–0.33)</td>
<td>0.44 (0.38–0.49)</td>
</tr>
<tr>
<td>IMRT/IMAT V95 = 98%</td>
<td>0.77 (0.70–0.84)</td>
<td>0.84 (0.78–0.89)</td>
</tr>
<tr>
<td>IMRT/IMAT V95 = 95%</td>
<td>0.68 (0.61–0.74)</td>
<td>0.78 (0.72–0.83)</td>
</tr>
</tbody>
</table>

Note: An upper CI limit below 1.00 represents significantly different odds ratios between techniques in the paired Monte Carlo test.
dose-response relationships. The practice change in many centers towards boosting only the tumor bed with a margin, rather than the whole posterior fossa, means that neurocognitive decline attributable to cerebellar irradiation would depend on the size and location of the tumor. Unfortunately, despite extensive research, the cerebellar contribution to cognitive and affective regulation remains poorly understood.

The potential risk of tumor relapse from hippocampal-sparing radiotherapy needs to be defined. However, the hippocampus and SVZ made up only 1.3% of the whole-brain volume on average for the patients in our study. Thus, only a small portion of the target is underdosed.

Our study extends the recent study by Redmond et al., in that we compare not only IMRT with standard opposing fields but also IMAT and IMPT. IMRT is generally used with caution in children because of concerns about secondary malignancies when exposing large areas to low doses of radiation. In proton therapy, the risk of developing radiation-induced cancers due to secondary neutron irradiation is of special concern in children. The IMPT plans in this study used spot-scanned delivery, which exposes the patient to considerably lower secondary neutron doses than passive scattering techniques. Although beyond the scope of this study, the risk of secondary malignancies needs to be considered in the choice of treatment modality, especially when addressing the spinal part of a CSI treatment course. Furthermore, Merchant et al. have stated that a reduction in low- and medium-dose volumes in the supratentorial brain benefits long-term cognitive outcome, which again favors the IMPT technique.

In summary, we demonstrate the dosimetric feasibility of sparing the hippocampus and SVZ during cranial irradiation, along with estimates of the potential clinical benefit. Our estimates show that the frequency of neurological adverse effects of radiotherapy could be considerably reduced, especially with intensity-modulated proton therapy. Validation of this strategy should come from large prospective clinical trials. Hopefully, our study can inspire such a trial, preferably with IMPT, because this technique is predicted to offer the greatest patient benefit.

Conflict of interest statement. P.M.R. has a research agreement with Varian Medical Systems. All other authors: None declared.

Acknowledgments
Malin Blomstrand and N. Patrik Brodin are co-first authors.

Funding
This work was supported by the Danish Child Cancer Foundation (to N.P.B.), Wilhelm and Martina Lundgren’s scientific fund 2 (grant vet 2-43/2010 ref 256 to M.B.), Sahlgrenska University Hospital’s foundations (grant 5889 act 8181 to M.B.), Jubileumsklinikken’s anti-cancer research fund (grant 2006:10 to M.B.), and National Cancer Institute (grant 2P30 CA 014520-34 to S.M.B.).

References
STUDY IV
Hippocampal sparing radiotherapy for pediatric medulloblastoma: impact of treatment margins and treatment technique

N. Patrik Brodin¹², Per Munck af Rosenschöld¹², Malin Blomstrand³, Anne Kiil-Berthlesen¹, Christian Hollensen¹, Ivan R. Vogelius¹, Birgitta Lannering⁴, Søren M. Bentzen⁵, Thomas Björk-Eriksson⁶

¹Radiation Medicine Research Center, Department of Radiation Oncology, Rigshospitalet, Copenhagen, Denmark
²Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
³Department of Oncology, Sahlgrenska University hospital, Gothenburg, Sweden, Centre for Brain Repair and Rehabilitation, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden
⁴Department of Pediatric Oncology, the Queen Silvia Children's Hospital, University of Gothenburg, Gothenburg, Sweden
⁵Department of Human Oncology, University of Wisconsin Medical School, Madison, WI, USA
⁶Department of Oncology, Skåne University Hospital and Lund University, Lund, Sweden

Funding

This work was supported by a grant from the Danish Child Cancer Foundation
Abstract

Purpose: Given the recent evidence for cognitive dysfunction due to hippocampal irradiation of pediatric patients, we assessed the prospect of reducing the risk of cognitive impairment using modern radiation therapy techniques for pediatric medulloblastoma (MB). Here, we investigated how varying treatment margin impacted the risk of neurocognitive impairment in relation to hippocampal sparing intensity-modulated radiotherapy (IMRT) and proton therapy (PT) when compared to 'current standard' 3D Conformal Radiotherapy (3D CRT).

Materials and methods: There can be considerable heterogeneity between MB patients with respect to tumor location relative to the hippocampal region. To reflect this variation, we included 17 pediatric MB patients in this study. Standard treatment plans were generated using current standard 3D CRT and using a hippocampal sparing approach with IMRT and spot-scanned PT. Three different treatment margins were used for conformal tumor bed boost. The risk of neurocognitive impairment was estimated based on dose-response models from long-term survivors of pediatric CNS malignancies and we compared risks between 3D CRT, hippocampal sparing IMRT and PT using conformal boost techniques with whole posterior fossa treatment.

Results: The mean dose to the hippocampus and the corresponding risks of cognitive impairment were decreased with decreasing treatment margins (p<0.05). The largest risk reduction, however, was seen when applying hippocampal sparing PT and the estimated risk of impaired task efficiency (95% confidence interval) was 92% (66-98%), 81% (51-95%) and 50% (30-70%) for 3D CRT, IMRT and PT for the smallest investigated boost margin and respectively 98% (78-100%), 90% (60-98%) and 70% (39-90%) if boosting the whole posterior fossa. We also found that the distance between the closest point of the planning target volume (PTV) and the center of the hippocampus (average between left and right) can be used to predict the mean hippocampal dose for the given treatment technique.

Conclusions: We estimated a considerable clinical benefit of applying a hippocampal sparing approach using modern radiotherapy techniques. In choosing treatment margins, the increase in target coverage with increasing margin size should be weighed against the increase in estimated risk of neurocognitive impairment quantified here.
Introduction

Medulloblastoma (MB) is the most commonly occurring malignant brain tumor in children, accounting for about 20% of all pediatric central nervous system (CNS) malignancies.\(^1\) Radiotherapy plays a vital role in the treatment of MB which given together with surgery and post-irradiation chemotherapy results in 5-year survival rates of 75-85% for standard-risk patients (without spinal metastases, cerebrospinal fluid involvement or large unresected tumor).\(^2\) Although effective, this aggressive treatment is associated with a considerable risk of late occurring side effects, of which neurocognitive decline is one of the most devastating and, unfortunately, common side effects.\(^1,3\) Irradiation of the CNS is a major contributor to this decline and a relationship between radiation dose to the brain and cognitive impairment has frequently been reported in studies of pediatric brain tumor patients.\(^4-9\)

Specifically, a strong correlation has been shown between dose to the hippocampus (referring to both the left and right hippocampus as one common structure) as well as the temporal lobe and cognitive outcome.\(^4,8,10\) The hippocampus, which is one of the major sites of neurogenesis in the adult brain of humans\(^11\) as well as rodents\(^12\), is situated within the temporal lobe. Neurogenesis is thought to be important for memory formation\(^13,14\) and cognitive deficits have been reported after radiation-induced impaired hippocampal neurogenesis in rodents.\(^15,16\) Hence, it appears reasonable to regard the hippocampus as the main risk organ for cognitive deficits after brain irradiation. Accordingly, reducing the hippocampal radiation dose would likely reduce the risk of late neurocognitive side-effects. We hypothesize that this can be achieved using modern inverse-planned radiotherapy techniques with dose-volume constraints set to limit the dose to the hippocampus while still treating the planning target volume (PTV) to a high dose. We previously demonstrated the feasibility of sparing the hippocampus and subventricular zone (SVZ) in whole-brain irradiation of MB, with little anatomical variation between patients.\(^17\) When considering the high dose boost however, the variation between individuals is considerably larger.

The radiotherapy target in MB has remained fairly constant with whole craniospinal irradiation (CSI) of 23.4-36 Gy and a high dose boost to the posterior fossa (PF) to 54-55.8 Gy. Many centers are now moving away from whole PF irradiation in favor of boosting only the primary tumor bed including a safety margin.\(^6,18-22\) This introduces some heterogeneity between patients in the target for the high dose boost, dependent on the extent and anatomical position of the primary tumor. The efficacy of a conformal tumor bed boost compared to treating the whole PF is currently being evaluated in the ongoing randomized ACNS 0331 trial. Although promising results have been reported from prospective cohort
studies applying a conformal tumor bed boost, there appears to be a lack of consensus regarding what margin should be added to the tumor bed (gross tumor volume (GTV)) to yield the clinical target volume (CTV). The GTV to CTV margins applied in these reports range from 0.5 to 1.5 cm, all with a further 0.5 cm margin to yield the corresponding PTV.

In this study we analyze how the size and position of the primary tumor bed as well as applying different treatment margins to the tumor bed will influence the dose to the hippocampus and whether a hippocampal sparing approach for the boost treatment is predicted to yield a clinically relevant reduction in dose. We estimate the risk of cognitive decline using published dose-response models for late neurocognitive impairment based on long-term follow-up of CNS patients in the childhood cancer survivor study. We compared conventional 3D conformal radiotherapy (3D CRT) with two hippocampal sparing techniques; intensity-modulated radiotherapy (IMRT) and spot-scanned proton therapy (PT).

Materials and methods

Patient material and treatment planning

Seventeen pediatric MB patients were included in this study, of which 13 were treated at Rigshospitalet in Copenhagen, Denmark in 2004-2009 and 4 were treated at Sahlgrenska University Hospital in Gothenburg, Sweden in 2002-2007. The median age at treatment of all 17 patients was 6 years (range: 4-15 years). A requirement for inclusion was access to pre-radiotherapy computed tomography (CT) and magnetic resonance imaging (MRI) scans for each patient. The tumor bed, including operation cavity, the PF and the hippocampus were delineated on fused transversal CT and T1-weighted MRI scans by an experienced neuroradiologist. The eyes, SVZ, cochlea, pituitary gland, parotids, optic chiasm, brainstem, and whole brain were also delineated as organs at risk (OARs).

To test the effect of applying different treatment margins, the CTVs were expanded around the pre-operative tumor bed with isotropic margins of 0.5, 1.0 and 1.5 cm while constraining the lateral, caudal and anterior extension of the CTV volumes to be within the PF. Cranially, the CTVs were limited to extending at most 0.5 cm outside of the PF; no constraint was applied in the posterior direction. The PTVs for each patient (henceforth referred to as PTV A, PTV B and PTV C) were expanded from the corresponding CTV volumes with a 0.5 cm isotropic margin. A PTV was also created with a 0.5 cm margin from the PF volume (PTV D) to allow comparison with the conventional boost treatment. Figure 1 shows an example of the CTV and PTV expansion for one of the patients.
Figure 1. The expansion of clinical target volume (CTV) and planning target volume (PTV) is shown in relation to the tumor bed on a transversal computed tomography image. This illustrates the difference in the resulting target volume for one patient based on which CTV margin is applied, with a CTV to PTV margin of 0.5 cm in all scenarios.

For each patient and for each of the four PTVs, individual treatment plans for the high dose boost were generated with 3D CRT, IMRT and PT, respectively (Eclipse v10, Varian Medical Systems, Palo Alto, USA). For the 3D CRT technique, two to four treatment fields were used, angled from the latero-posterior direction (two opposed and two oblique) with the aim of covering the PTV to the best possible extent. Five fields angled from the posterior direction were used for the inverse-planned IMRT technique, where the main objective was to spare the hippocampus while still achieving a good PTV coverage and clinically acceptable doses to the remaining OARs. The same was done for the spot-scanned PT technique but with only a single treatment field from the posterior direction. The different treatment techniques are illustrated for one of the patients in Figure 2.
Figure 2. An example for one patient of the field setup for the different treatment techniques and the resulting dose distributions in color-wash. The target, which in this example is the PTV B, is shown in cyan and the hippocampus is the magenta-colored structure.

The trade-off between sparing the hippocampus and covering the PTV is controlled by prioritizing between the two through a set of dose-volume objectives assigned in the treatment optimization. We started each treatment optimization at a given set of dose-volume objectives; presented in Table 1. The objectives for the OARs were set at a level where the optimizer would be pressed to fulfill them for the larger target volumes (PTV C and PTV D) and if an objective was met it was lowered to further reduce the dose to that structure. The aim of this treatment optimization strategy was to minimize user-dependence and thus increase the reproducibility of our results.
Table 1. Dose-volume objectives used for the inverse-planned treatment techniques.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Volume (%)</th>
<th>Dose (Gy)</th>
<th>Priority</th>
<th>Volume (%)</th>
<th>Dose (Gy)</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IMRT</td>
<td></td>
<td>PT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTV</td>
<td>0</td>
<td>31.1</td>
<td>240</td>
<td>0</td>
<td>31.1</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>30.1</td>
<td>240</td>
<td>100</td>
<td>30.1</td>
<td>240</td>
</tr>
<tr>
<td>CTV</td>
<td>0</td>
<td>31.1</td>
<td>240</td>
<td>0</td>
<td>31.1</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>30.1</td>
<td>240</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hippocampus</td>
<td>26</td>
<td>9.0</td>
<td>80</td>
<td>26</td>
<td>9.0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>12.0</td>
<td>80</td>
<td>0</td>
<td>12.0</td>
<td>100</td>
</tr>
<tr>
<td>Brainstem</td>
<td>0</td>
<td>28.1</td>
<td>75</td>
<td>0</td>
<td>28.5</td>
<td>75</td>
</tr>
<tr>
<td>Optic chiasm</td>
<td>0</td>
<td>15.7</td>
<td>50</td>
<td>0</td>
<td>15.7</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>19.0</td>
<td>65</td>
<td>13</td>
<td>17.5</td>
<td>65</td>
</tr>
<tr>
<td>Cochlea</td>
<td>50</td>
<td>15.2</td>
<td>65</td>
<td>50</td>
<td>14.4</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>17.2</td>
<td>65</td>
<td>31</td>
<td>16.0</td>
<td>65</td>
</tr>
<tr>
<td>Eyes</td>
<td>27</td>
<td>1.8</td>
<td>65</td>
<td>24</td>
<td>0.3</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.7</td>
<td>65</td>
<td>2</td>
<td>0.5</td>
<td>65</td>
</tr>
<tr>
<td>Parotids</td>
<td>26</td>
<td>4.9</td>
<td>60</td>
<td>23</td>
<td>2.6</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>5.8</td>
<td>60</td>
<td>5</td>
<td>4.3</td>
<td>60</td>
</tr>
<tr>
<td>Pituitary gland</td>
<td>0</td>
<td>14.2</td>
<td>50</td>
<td>0</td>
<td>13.6</td>
<td>50</td>
</tr>
</tbody>
</table>
Analyzing dosimetric data and the dependence on target size and position

The calculated doses and delineated structures for each patient and treatment plan were exported as DICOM-files and analyzed using the computational environment for radiotherapy research (CERR) software to create corresponding dose-volume histograms (DVHs). Extraction of doses to target and OARs, as well as the corresponding volumes, was carried out based on the DVHs using the MATLAB software (The MathWorks, Inc.).

The total prescribed dose was 54 Gy to the boost target, planned as 23.4 Gy from the CSI and 30.6 Gy from the boost treatment. In a previous analysis focused on hippocampal sparing during the whole-brain part of CSI, we derived dosimetric relationships between hippocampal dose and the percentage of the whole-brain that receives at least 95% of the prescribed dose (V95). The hippocampal dose contribution, stemming from the whole-brain irradiation was taken as the dose that corresponds to 95% of the whole-brain receiving at least 95% of the prescribed dose for 3D CRT, IMRT and PT, respectively and was added to the hippocampal dose from the boost treatment. We included the uncertainty in the CSI hippocampal dose contribution by including the range of estimated doses among the patients in the previous analysis.

We tested for a correlation between the size of the boost PTV and the absorbed dose to the hippocampus. We also tested whether a simple metric describing the distance or overlap between PTV and hippocampus could be used as a surrogate to predict the hippocampal dose from the boost treatment.

Estimating the neurocognitive impairment risk for the different treatment techniques and margins

We used recently published dose-response models for estimating the long-term risk of impaired task efficiency, organization and memory to estimate the risk of neurocognitive impairment. These logistic dose-response models were derived based on cognitive outcome data from long-term survivors of childhood CNS malignancies, stratified into a group with only MB/PNET patients. The models stem from the relationship between temporal lobe dose and the corresponding type of cognitive impairment. In the current analysis we make the ansatz that the dose-risk estimates for the hippocampus alone are the same as the dose-risk estimates for the entire temporal lobe. This corresponds to assuming that the hippocampus is the key structure for radiation-induced neurocognitive decline following temporal lobe irradiation. The dose-response models are given by the following relation:
where $D$ is the absorbed dose in Gy, $OR_{10}$ is the odds ratio (OR) for the given impairment at 10 Gy, $p_0$ is the baseline risk of impairment at zero dose and $p_D$ is the risk of impairment at dose $D$. The baseline risks were taken from patients in the CNS survivor cohort that did not receive any temporal lobe irradiation. The given $OR_{10}$ with 95% confidence intervals (CIs) are 2.95 (1.66-5.22), 2.21 (1.04-4.70) and 1.45 (0.91-2.30) for impaired task efficiency, organization and memory, respectively. The corresponding baseline risks ($p_0$) are 24.0%, 12.3% and 24.6%. The mean hippocampal dose from each of the different treatment plans was entered along with the different ORs into Equation 1 to estimate the risk of the respective neurocognitive impairment.

Statistical analysis

For the estimated risks of cognitive impairment, the largest uncertainty stems from the ORs used in the dose-response models which are given with 95% CIs. For each of the neurocognitive endpoints, a Monte Carlo (MC) sampling technique was applied by randomly drawing 10,000 samples from log-normal distributions matching the mean and 95% CI of the $OR_{10}$ dose-response parameters. We then derived the mean and 95% CI of the estimated risk of impairment in each of the samples. Finally the 95% CI was extracted as the 2.5 - 97.5 percentile of the 10,000 risk estimate values, averaged over the patient group.

To test whether the estimated risk of impairment was significantly different between treatment techniques, the difference in risk between two treatment techniques was extracted for each of the 10,000 MC samples. The mean and 95% CI of risk difference for the whole patient group was then calculated by inverse variance weighting of the difference estimates for each patient. To account for potential underestimation of the variance due to the relatively small number of patients, this was done using a bootstrapping procedure. Here, 100,000 samples of the 17 patients were drawn with replacement. A point estimate and 95% CI of the mean risk difference was derived for each sample as explained above. A normal distribution was matched to each of these point estimates and CI. One sample was then randomly drawn.
from each of the 100,000 distributions, giving the final risk difference between treatment techniques with 95% CI as the 2.5 - 97.5 percentile of these randomly drawn samples.

The association between PTV size and hippocampal dose was estimated using Spearman’s rank correlation coefficient (Spearman’s $\rho$). Whether these correlations were statistically significant was tested by bootstrapping over the data for the 17 patients. For each correlation analysis, 1,000,000 bootstrap samples were drawn with replacement and the corresponding Spearman’s $\rho$ calculated for each sample. The 95% CIs of the correlation coefficients were taken as the 2.5 - 97.5 percentile of the randomly sampled Spearman’s $\rho$ coefficients. The correlations were considered statistically significant if the 95% CI of Spearman’s $\rho$ did not cross zero. Corresponding one-sided p-values were estimated as the proportion of sampled correlation coefficients that did cross zero in the direction (negative or positive) opposite of the point estimate.

**Results**

The resulting tumor bed and target volumes for the different treatment margins are presented in Table 2.

**Table 2.** Median target volumes (range) for the 17 patients.

<table>
<thead>
<tr>
<th>Target structure</th>
<th>Tumor bed (GTV)</th>
<th>CTV 0.5 cm</th>
<th>CTV 1.0 cm</th>
<th>CTV 1.5 cm</th>
<th>CTV PF</th>
<th>PTV A</th>
<th>PTV B</th>
<th>PTV C</th>
<th>PTV D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume (cm$^3$)</td>
<td>17.2 (2.9-60.9)</td>
<td>42.0 (12.0-107.4)</td>
<td>71.3 (30.5-131.8)</td>
<td>107.1 (57.3-185.1)</td>
<td>161.0 (115.2-267.0)</td>
<td>81.9 (30.0-187.3)</td>
<td>127.8 (62.1-224.7)</td>
<td>177.8 (104.6-261.3)</td>
<td>276.2 (213.2-420.2)</td>
</tr>
</tbody>
</table>

Figure 3 shows the spatial distribution of delineated tumor bed volumes and hippocampi between patients.
**Figure 3.** The spatial distribution of delineated primary tumor beds in relation to the hippocampi for the patients in our cohort. The tumor beds are shown in red and the hippocampi in blue with the semi-transparent structures for all patients overlaid on each other. The locations of these structures are overlaid on a single posterior fossa (PF) contour representative of most PF volumes in our patient group. For graphical clarity, the tumor beds for two patients with very irregular size and outlying position of their operation cavities were excluded from this illustration (these were, however, included in all quantitative analyses).

The resulting hippocampal doses for each treatment technique and target volume are shown in Table 3, with and without the dose contribution from the whole-brain part of treatment. The mean doses from the boost treatment to all other OARs included in this study are also given in Table 3. Here, it is clearly shown that a larger boost target volume results in higher dose to the hippocampus. Both of the hippocampal sparing techniques result in a substantial reduction in mean dose to the hippocampus. For the remaining OARs, PT resulted in lower doses across the board compared to the photon techniques (p<0.003 for all OARs and margins in the Wilcoxon signed rank test). The IMRT plans were superior compared to 3D CRT for sparing the cochlea (p<0.001), parotids (p<0.002) and whole-brain (although not statistically significant with p>0.1 for all margins) while resulting in higher doses to the optic chiasm (p<0.05 for all target margins except the PF), pituitary gland (p<0.045) and the eyes (p<0.003). The SVZ received virtually no dose from the boost regardless of treatment technique and margin.
Table 3. Mean doses (range) from the boost treatment of 30.6 Gy (from 23.4 up to 54 Gy) for each treatment technique and margin.

<table>
<thead>
<tr>
<th>Treatment technique</th>
<th>Target/ risk organ</th>
<th>3D CRT</th>
<th>IMRT</th>
<th>PT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.4-25.5)</td>
<td>(0.7-28.1)</td>
<td>(1.1-28.9)</td>
<td>(17.0-29.4)</td>
</tr>
<tr>
<td></td>
<td>Hippocampus, boost dose (Gy)</td>
<td>14.3</td>
<td>18.0</td>
<td>20.6</td>
</tr>
<tr>
<td></td>
<td>Hippocampus, total dose* (Gy)</td>
<td>36.6</td>
<td>40.2</td>
<td>42.8</td>
</tr>
<tr>
<td></td>
<td>Cochlea (Gy)</td>
<td>17.9</td>
<td>22.0</td>
<td>24.9</td>
</tr>
<tr>
<td></td>
<td>Pituitary gland (Gy)</td>
<td>2.9</td>
<td>5.9</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>Eyes (Gy)</td>
<td>0.6</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>Parotids (Gy)</td>
<td>6.1</td>
<td>8.2</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>Optic chiasm (Gy)</td>
<td>2.5</td>
<td>4.2</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>Whole brain (Gy)</td>
<td>6.1</td>
<td>7.1</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>SNZ (Gy)</td>
<td>0.4</td>
<td>0.5</td>
<td>0.7</td>
</tr>
</tbody>
</table>

*Hippocampal dose from 23.4 Gy whole-brain irradiation and boost treatment of 30.6 Gy.
Figure 4 shows the distribution of hippocampal doses between patients for the different boost treatment scenarios, illustrating an approximately linear increase in dose with treatment margin.

**Figure 4.** Distribution of mean hippocampal dose among the 17 patients in this study presented in a box-and-whiskers plot for the different treatment techniques and target volumes. The red horizontal bar shows the median, the box edges the 25th and 75th percentiles and the whiskers show the range of mean doses. Outliers are plotted as red plus signs.

There were significant correlations between the size of the PTV in cm³ and the mean hippocampal dose with a Spearman’s ρ of 0.61 (95% CI: 0.33-0.85) for 3D CRT, 0.62 (95% CI: 0.34-0.84) for IMRT and 0.59 (95% CI: 0.31-0.83) for PT, all with one-sided p-values<0.0001. The size of the PTV is however not an optimal measure to predict the hippocampal dose since it does not consider how the PTV is positioned in relation to the hippocampus. We found that a simple metric well suited for predicting the hippocampal dose in our patient group was the distance from the closest point of the PTV to the hippocampal center of mass (taken as the average distance to the center of mass of the left and right part of the hippocampus).
For distances up to 2 cm we fitted linear models with 95% prediction bounds describing the relation between this distance and the mean hippocampus dose as shown in Figure 5. The same linear relation does not hold for larger distances, since the lowest achievable hippocampal dose, for the respective treatment technique, appears to be reached. This seems to occur at a distance of about 2 cm for PT, 2.5 cm for IMRT and somewhere between 3-4 cm for 3D CRT, suggesting that if the distance between the closest PTV point and the hippocampus center is larger than this there is no need to apply hippocampal sparing.

Figure 5. The relation between mean hippocampus dose from the boost treatment and the distance between closest point of the planning target volume (PTV) and hippocampus center of mass is shown for the three different treatment techniques. The 95% prediction bounds give the 95% certainty for estimating the hippocampal dose based on this distance.

The estimated risks of neurocognitive impairment, based on the mean hippocampal dose from the boost and whole-brain treatment, are shown in Figure 6. The estimated risks are considerably reduced for PT
compared to the photon techniques. Although the 95% CIs are quite large, the estimated risks with PT were significantly lower for all treatment margins in a paired statistical comparison (cf. Table 4). There was also a significant difference between IMRT and 3D CRT in estimated impaired organization and task efficiency. There is also a difference between treatment margins which, again, appears to be linearly increasing when moving to larger treatment margins. There is thus a potential of reducing the risk of neurocognitive impairment through smaller target margins for the high dose boost, although the most substantial risk reduction is estimated to come from applying hippocampal sparing.

Figure 6. The estimated risk of impaired task efficiency, organization and memory are presented for each treatment technique and margin for a craniospinal dose of 23.4 Gy and a high dose boost up to 54 Gy.
The vertical bars show the 95% confidence intervals estimated through 10,000 Monte Carlo samples over the uncertainty in the corresponding dose-response parameters.

**Table 4.** Bootstrapped risk differences (95% confidence intervals (CIs)) between treatment techniques. Differences are statistically significant at the 5% level if the lower limit of the 95% CI does not cross zero.

<table>
<thead>
<tr>
<th>Risk difference</th>
<th>3D CRT – IMRT</th>
<th>IMRT – PT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target / risk organ</strong></td>
<td><strong>PTV A</strong></td>
<td><strong>PTV B</strong></td>
</tr>
<tr>
<td>Task efficiency (%)</td>
<td>11 (4-19)</td>
<td>9 (3-17)</td>
</tr>
<tr>
<td>Organization (%)</td>
<td>21 (5-36)</td>
<td>21 (4-36)</td>
</tr>
<tr>
<td>Memory (%)</td>
<td>11 (-1-22)</td>
<td>12 (-1-23)</td>
</tr>
</tbody>
</table>

**Discussion**

To the best of our knowledge, this is the first study to perform a detailed investigation of the radiation dose to the hippocampus from the combined course of RT including varying margins for the high-dose boost in MB treatment. The effect on hippocampal dose from varying treatment margins was an approximately linear increase in dose with increasing target margin. This did, however, also depend on the location of the primary tumor bed.

Although the size of the tumor bed target for the high-dose boost varied considerably among the 17 patients in our study, most tumors were located close to the central axis of the posterior cranial fossa. Most noteworthy is that the reduction in hippocampal dose from using smaller treatment volumes is modest compared to applying a hippocampal sparing technique.

Here, we consider the dose contribution from both the whole-brain part of CSI and from the high-dose boost. As expected, the relative dose contribution to the hippocampus is mainly from the whole-brain irradiation, cf. Table 3, although this also depends on the size of boost margins as well as how much of
the high-dose boost that is prescribed from the CSI versus the boost treatment. The results presented here can, however, be extrapolated to alternative MB treatment regimens.

Performing hippocampal sparing requires access to high-resolution MRI scans fused with CT images for radiation treatment planning. Contouring the hippocampus and other intra-cranial structures may require an experienced neuro-radiologist. The RTOG trial 0933 for hippocampal sparing of patients with brain metastases does, however, provide a hippocampal contouring atlas, which may aid in delineation.

We also estimated a considerable clinical benefit of applying a hippocampal sparing approach using modern radiotherapy techniques. The estimated risks of neurocognitive impairment are based on long-term follow-up data from 818 survivors of pediatric CNS malignancies, presented by Armstrong et al.4 Our estimates are thus subject to the statistical uncertainty in the dose-response parameters as well as the uncertainty in data collection through neurocognitive testing questionnaires and retrospective dosimetry applied in their study. Also, we assume that sparing the hippocampus is equivalent to sparing of the whole temporal lobe in the Armstrong study with respect to neurocognitive functioning. This assumption is further strengthened by a recent prospective study by Redmond et al, who showed a significant correlation of decreased motor speed and dexterity with mean dose to the hippocampus as well as the temporal lobe.10 This would suggest that mean hippocampal dose is an appropriate metric for estimating neurocognitive impairment although this should be validated in a larger patient material.

Redmond et al. did not see a significant correlation between neurocognitive outcome and dose to the SVZ consistent with the unclear role of SVZ irradiation in this setting. Hence we did not consider it as a cognitive risk organ in this study. In addition to hippocampal irradiation, the tumor itself, as well as surgery and aggressive chemotherapy, are also likely contributors. Irradiating the cerebellum to high doses might also add to the risk of cognitive decline as this could interrupt supratentorial connections between the cerebellum and frontal part of the brain, as suggested by Armstrong et al.4 Although this may also be attributed the surgical resection, it may be a further argument for applying a conformal tumor bed boost, rather than boosting the whole PF. The cerebellar contribution to higher cognitive functioning is, however, still poorly understood making it difficult to factor in such an effect. Irradiating the prefrontal cortex in the whole-brain treatment likely also plays a role in the cognitive decline of MB patients.

We showed that it may be possible to predict the mean hippocampal dose based on the distance between the closest point of the PTV to the center of the hippocampus (average between left and right). Measuring this distance for a MB patient before engaging in time consuming treatment planning can provide estimates of the hippocampal dose and corresponding risk of cognitive impairment depending on the
choice of treatment technique. This simple metric could thus help the radiation oncologist deduce whether the patient should receive the boost treatment with PT, which is not yet widely available, and/or whether the patient would benefit from hippocampal sparing radiotherapy.

The importance of limiting the dose to the hippocampus in pediatric patients receiving cranial irradiation is becoming increasingly evident. Here, we present a detailed analysis of the dose to the hippocampus and corresponding risks of neurocognitive impairment from the high-dose boost in MB treatment. Proton therapy shows most promise for hippocampal sparing due to the very sharp dose gradients achievable through the physical differences in energy deposition compared to photons. Availability of PT is still limited and here we showed that considerable hippocampal sparing is also possible using IMRT which makes this an attractive alternative option. In this analysis, we predict a correlation of the cognitive outcome with the selection of treatment margins. Specifically, treating with smaller margins from the tumor bed to the PTV compared to whole posterior fossa irradiation is estimated to reduce the risk of cognitive decline. The decrease in risk of neurocognitive impairment estimated must, however, be balanced against reducing the dose inside parts of the classical target volume. Although the hippocampus corresponds to only ~1% of the whole-brain volume limiting the dose here could, potentially, increase the risk of tumor recurrence. Clearly, hippocampal sparing should therefore be pursued within the setting of a prospective clinical trial in order to validate its effectiveness and safety.
Reference List


(9) Mulhern RK, Kepner JL, Thomas PR, Armstrong FD, Friedman HS, Kun LE. Neuropsychologic functioning of survivors of childhood medulloblastoma randomized to receive conventional or


STUDY V
Modeling freedom from progression for standard-risk medulloblastoma: A mathematical tumor control model with multiple modes of failure

N. Patrik Brodin\textsuperscript{1,2}, Ivan R. Vogelius\textsuperscript{1}, Thomas Björk-Eriksson\textsuperscript{3}, Per Munck af Rosenschöld\textsuperscript{1,2}, Søren M. Bentzen\textsuperscript{1,4}

\textsuperscript{1}Radiation Medicine Research Center, Department of Radiation Oncology, Rigshospitalet, Copenhagen, Denmark

\textsuperscript{2}Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark

\textsuperscript{3}Department of Oncology, Skåne University Hospital and Lund University, Lund, Sweden

\textsuperscript{4}Department of Human Oncology, University of Wisconsin Medical School, Madison, WI, USA

\textbf{Funding}

Danish Child Cancer Foundation to N.P.B.; National Cancer Institute (grant no. 2P30 CA 014520-34) to S.M.B.

\textbf{No conflicts of interest in relation to the submitted work}

\textbf{Running title}

Tumor control model for standard-risk medulloblastoma
Abstract

Background: As pediatric medulloblastoma (MB) is a relatively rare disease, it is important to extract the maximum information from trials and cohort studies. Here, a framework is developed for modeling tumor control with multiple modes of failure and time-to-progression for standard-risk MB, using published pattern of failure data.

Methods: Outcome data for standard-risk MB published after 1990 with pattern of relapse information were used to fit a tumor control dose-response model addressing failures in both the high-dose boost volume and the elective craniospinal volume. Estimates of 5-year event-free survival from two large randomized MB trials were used to model the time-to-progression distribution. Uncertainty in freedom from progression (FFP) was estimated by Monte Carlo sampling over the statistical uncertainty in input data.

Results: The estimated 5-year FFP (95% confidence intervals) for craniospinal doses of zero, 18, 24 and 36 Gy while maintaining 54 Gy to the posterior fossa was 76% (0-81%), 78% (73-81%), 79% (76-82%) and 80% (77-84%) respectively. The uncertainty in FFP was considerably larger for craniospinal doses below 18 Gy, reflecting the lack of data in the lower dose range.

Conclusions: Estimates of tumor control and time-to-progression for standard-risk MB provides a data-driven setting for hypothesis generation or power calculations for prospective trials. The presented methods can also be applied to incorporate further risk-stratification for example based on molecular biomarkers, when the necessary data become available. The result of this modeling study is a rather shallow dose-response model for MB which is consistent with the negative result of the recent PNET-4 trial.
**Introduction**

Childhood medulloblastoma (MB) is fortunately a relatively rare disease. Consequently, performing randomized trials is a major challenge requiring large collaborative networks with many participating centers (1-3). Thus, it is important to extract the maximum information from the combination of randomized trials, prospective single arm studies and retrospective analyses.

The radiotherapy target volumes for standard-risk MB patients has remained fairly consistent over recent decades where treatment has mainly been improved through scheduling of immediate radiotherapy and post-radiation chemotherapy along with reducing the craniospinal irradiation (CSI) dose from 36 to 23.4 Gy (4, 5). There have been further attempts to lower the CSI dose to 18 Gy (6-8), albeit with a very limited number of patients. CSI is generally followed by a high-dose boost to the posterior fossa (PF) up to 54-55.8 Gy. Currently, the tendency is to boost the pre-operative tumor bed with a 1-2 cm margin rather than the whole PF, hypothesizing that this will yield the same local control (4). The results of the ongoing ACNS 0331 trial, randomizing MB patients to 23.4 Gy CSI vs. 18 Gy CSI and tumor bed boost vs. whole PF boost will provide further insight into correct dose and target volume in this disease.

MB patients are at risk of two different *modes of failure* after radiotherapy; 1) failing in the primary site and 2) failing in the neuraxis from subclinical spread. Consequently, where traditional dose-response studies rely on relapse rate, a model of tumor control for pediatric MB should consider the boost and elective (craniospinal) dose and risk of recurrence separately. Ideally, time-to-event should be taken into account, so that freedom from progression (FFP) over time can be estimated as a function of radiation dose.

MB has recently been characterized as encompassing four distinct molecular subgroups (9). Complementing the classical clinical risk variables in MB with stratification according to these molecular subgroups results in substantial difference in overall survival between groups, far more prognostic than stratifying solely on clinical information (9-11).

The current paper derives a dose-response model for FFP with two modes of failure from previously published data. Actuarial data on FFP is used to combine the dose-response model with time to progression data in a mixture model, resulting in estimates of the FFP as function of
follow-up time for given radiation doses to the boost and elective regions. The FFP curve that can thus be compared directly with the Kaplan-Meier curves from clinical outcome studies. As an example, we use the model to derive a post-hoc power estimate for the recently published PNET-4 trial comparing two different radiotherapy fractionation schemes (1). This paper also discusses the potential development of dose-response models for different molecular subgroups of MB. The necessary data can likely be extracted from secondary analysis of the PNET-4 and ACNS 0331 trials. Finally, we discuss how the model can be applied in clinical trial design and in silico simulations of altered treatment regimens.

**Methods and materials**

*Search strategy and inclusion criteria*

Pattern of failure information was collected, if available, from all outcome studies in standard-risk MB published after 1990. Available published data was found by searching the Pubmed database using the search string “radiotherapy AND (medulloblastoma) NOT (recurrent OR recurrence) NOT neuroblastoma NOT Ewing NOT glioblastoma NOT review[publication type] NOT infant” including limits (children 0-18, English language and publication date from January 1st 1990 to January 31st 2012). Additional studies were identified through manually screening reference lists and selected reviews. The selection process for the included studies is depicted in Figure 1. Studies without chemotherapy or where pattern of failure at 5 years post diagnosis could not be extracted were not included in the quantitative dose-response analysis.
**Figure 1.** Flow chart illustrating the selection process for papers included in the dose-response analysis.

*Tumor control with multiple modes of failure*

Table 1 lists all the studies that were included to derive the 5-year FFP dose-response model. Each study and dose level was weighted by the inverse variance of their 5-year Kaplan-Meier estimate of event-free survival (EFS).
Table 1. Pattern of failure data included in the analysis.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study design</th>
<th>5-year EFS (±SE)</th>
<th># patients</th>
<th>Boost dose (Gy)*</th>
<th>CSI dose (Gy)*</th>
<th>relapses boost site</th>
<th>relapses elective</th>
<th>relapses both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lannering et al. 2012 (1)</td>
<td>Randomized trial</td>
<td>78±3%</td>
<td>169</td>
<td>54.9</td>
<td>23.0</td>
<td>4</td>
<td>25</td>
<td>6</td>
</tr>
<tr>
<td>Standard arm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperfractionated arm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Packer et al. 2006 (2)</td>
<td>Randomized trial</td>
<td>81±2.1%</td>
<td>379</td>
<td>54.1</td>
<td>23.0</td>
<td>20</td>
<td>24</td>
<td>16</td>
</tr>
<tr>
<td>Jakacki et al. 2004 (7)</td>
<td>Prospective cohort study</td>
<td>57±19%</td>
<td>7</td>
<td>53.1</td>
<td>17.7</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Goldwein et al. 1996 (6)</td>
<td>Prospective cohort study</td>
<td>70±20%</td>
<td>10</td>
<td>51.6</td>
<td>17.7</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Yasuda et al. 2008 (8)</td>
<td>Prospective cohort study</td>
<td>75±15%</td>
<td>8</td>
<td>47.2</td>
<td>17.7</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Taylor et al. 2003 (3)</td>
<td>Randomized trial</td>
<td>74.2±4.8%</td>
<td>90††</td>
<td>54.1</td>
<td>34.4</td>
<td>5</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Packer et al. 1999 (12)</td>
<td>Prospective cohort study</td>
<td>79±7%</td>
<td>65</td>
<td>54.1</td>
<td>23.0</td>
<td>2</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Oyharcabal-Bourden et al. 2005 (13)</td>
<td>Prospective cohort study</td>
<td>71.8±10.5%</td>
<td>71††</td>
<td>53.1</td>
<td>24.6</td>
<td>2</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Merchant et al. 2008 (4)</td>
<td>Prospective cohort study</td>
<td>83±5.3%</td>
<td>86</td>
<td>54.1††</td>
<td>23.0</td>
<td>3</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Douglas et al. 2004 (14)</td>
<td>Prospective cohort study</td>
<td>86±6.4%</td>
<td>33</td>
<td>54.1</td>
<td>23.0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Paulino et al. 2011 (15)</td>
<td>Prospective cohort study</td>
<td>75.1±7.6%</td>
<td>33†</td>
<td>54.5††</td>
<td>23.0</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Gentet et al. 1995 (16)</td>
<td>Prospective cohort study</td>
<td>74±21%</td>
<td>31</td>
<td>53.1</td>
<td>31.0†#</td>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

* Given as equivalent dose in 2 Gy fractions (EQD2) assuming $\alpha/\beta=10$
†† Posterior fossa given 55 Gy, further boost to tumor bed up to 62.3 Gy
§ Patients in the chemotherapy + radiotherapy arm
††† Out of 136 patients correctly classified, relative distribution of relapses assumed same as for the whole cohort
†† Posterior fossa given 35.4 Gy, further boost to tumor bed up to the given dose
* Out of 50 patients in total, the given EFS and pattern of failure data relates to the 33 standard-risk patients
# Whole brain treated to 27 Gy and spinal axis treated to 36 Gy

The control rates for the elective craniospinal volume and the PF boost volume were estimated for EQD2$_{10}$ (equieffective dose in 2 Gy fractions with $\alpha/\beta=10$ Gy) using the patterns of failure data from all included studies, see Table 1 for details. Further mention of radiotherapy doses data will refer to the standard scheme of 1.8 Gy per fraction.
Deriving dose-response parameters and handling synchronous failures.

Dose-response relations for the respective sites were derived as logistic functions with the logarithm of dose as a covariate, as presented in Bentzen and Tucker (17). Dose-response model parameters were fitted to the pattern of failure data in Table 1. The combined tumor control probability (TCP) is estimated as the product of TCPs for each target site, \( j \), i.e. assuming statistical independence between the \( R \) types of failures:

\[
TCP = \prod_{j=1}^{R} TCP_j = \prod_{j=1}^{R} \left( \frac{1 - P_{0,j}}{1 + \left( \frac{D_{50,j}}{D_j} \right)^{4\gamma_{50,j}}} + P_{0,j} \right)
\]

where \( P_{0} \) is the TCP with no radiation, \( D_{50,j} \) is the dose required to achieve an EFS of \((1 + P_{0,j})/2\), \( \gamma_{50} \) is the normalized dose-response gradient at the 50% control level. The \( \gamma_{50} \) may be interpreted as the increase in percentage points for a 1% relative increase in radiation dose. Here, the value of \( \gamma_{50,1} = \gamma_{50,2} \) (assumed as the same for the boost = 1 and elective = 2 volume) was estimated at 0.36 from the study by Thomas et al. comparing 5-year EFS after CSI doses of either 23.4 Gy or 36 Gy, without the use of chemotherapy (5).

Time-dependent FFP: Parametric mixture model

A parametric mixture model (18) using a log-logistic survival function was fitted to the Kaplan-Meier EFS curves in two large clinical MB studies (1, 2) published by Packer et al. in 2006 and the PNET-4 study published by Lannering et al. in 2012. A combined model was derived, weighted by the number of patients in each study, shown in Figure 2. In the mixture model, a proportion of the patients, \( P_{\text{cured}} \), are considered to be cured whereas the remainder, \( 1 - P_{\text{cured}} \), eventually fail, with the failure time distribution given by the log-logistic model fitted to outcome data. Thus, FFP and the corresponding hazard of relapse, \( h_{\text{relapse}} \) are given by;
Where, \( \alpha \) and \( \lambda \) are log-logistic model parameters derived from fitting the model to data and \( t \) is the observation time.

Here, \( \alpha \) and \( \lambda \) were estimated from the model fit and \( P_\perp \) was determined for each TCP estimate by setting the FFP at 5 years equal to the estimate from the TCP model:

\[
FFP(t) = (1 - P_\perp) \frac{1}{1 + \left( \frac{t}{\lambda} \right)^\alpha} + P_\perp \tag{2}
\]

\[
h_{relapse}(t) = -\frac{\partial \ln(FFP(t))}{\partial t} = -\frac{(P_\perp - 1)\alpha}{t^{1-\alpha} \lambda^\alpha + 2t + t^{\alpha+1} \lambda^{-\alpha}} \tag{3}
\]

The presented model has difficulty describing the time-to-failure distribution for very low TCP values where it yields negative values of \( P_\perp \) which are unphysical. Thus, when the model returns a negative \( P_\perp \) value this is forced to zero and the model altered by changing the \( \lambda \) parameter accordingly, giving a FFP model that flattens out earlier compared to the case shown in Figure 2.
Figure 2. Modeled freedom from progression fitting the log-logistic mixture model to outcome data. The resulting parameters for the combined model were $\alpha = 1.99$ and $\lambda = 2.39$ y$^{-1}$. The exponential model is included to illustrate that a one parameter model does not fit the data.

The probability density function of relapsing at a certain time point is:

$$P_{\text{relapse}}(t) = FFP(t) \cdot h_{\text{relapse}}(t)$$

(5)

Thus, changes in radiation dose-distribution directly influence $P_{\text{relapse}}$ which drives the change in FFP as a function of time.

*Synchronous failures vs. the assumption of independence*

To handle synchronous failures, the risks of failing in the elective and boost sites are assumed independent. However, this would produce a very low probability of synchronous failures that
are not in agreement with data, cf. Table 1. Here, we assume that the excess risk of synchronous failures is attributed failures in the primary site having spread to the elective site before a relapse is diagnosed. In a sensitivity analysis (cf. Appendix A in supplementary material), we showed that the model is not noticeably affected by relaxing this assumption. We fix the total recurrence rate to the observed rate, but the proportion of recurrences recorded as synchronous are re-defined to occur in the primary site (most often) or elective site (rare), such that the assumption of independence is met.

More precisely:

\[ \rho_{\text{elective}} + \rho_{\text{boost}} = \rho_{\text{elective}}^\text{true} + \rho_{\text{boost}}^\text{true} + \rho_{\text{sync}}^\text{true} \]

where the frequencies, \( \rho_{x}^\text{true} \) can be synchronous and the recorded failure frequencies, \( \rho_{x}^\text{recorded} \) in the elective, boost or synchronous site are subject to misclassifications because of potential spread before the recurrence is diagnosed. Given this, the following equations fully specify the problem of identifying the true recurrence origin:

\[ \rho_{\text{elective}}^\text{true} \cdot \rho_{\text{boost}}^\text{true} = \rho_{\text{sync}}^\text{true} \quad \text{(Assumption of independence)} \]

\[ \rho_{\text{elective}}^\text{true} + \rho_{\text{boost}}^\text{true} = \rho_{\text{elective}}^\text{recorded} + \rho_{\text{boost}}^\text{recorded} + \rho_{\text{sync}}^\text{true} + \rho_{\text{sync}}^\text{recorded} \quad \text{(where \( \rho_{x}^\text{true} \) includes synchronous)} \]

The true boost and elective frequencies are defined by:

\[ \rho_{\text{elective}}^\text{true} = \rho_{\text{elective}}^\text{recorded} + \rho_{\text{sync}}^\text{true} \]

\[ \rho_{\text{boost}}^\text{true} = \rho_{\text{boost}}^\text{recorded} + \rho_{\text{sync}}^\text{recorded} \quad \text{(boost site failures are assumed to consist of both true synchronous failures and those arising from the boost site although recorded as synchronous)} \]

This gives:

\[ \left( \rho_{\text{elective}}^\text{recorded} + \rho_{\text{sync}}^\text{true} \right) \left( \rho_{\text{boost}}^\text{recorded} + \rho_{\text{sync}}^\text{recorded} \right) = \rho_{\text{sync}}^\text{true} \]

\[ \rho_{\text{sync}}^\text{true} = \frac{\rho_{\text{elective}}^\text{recorded} \cdot \left( \rho_{\text{boost}}^\text{recorded} + \rho_{\text{sync}}^\text{recorded} \right)}{1 - \rho_{\text{boost}}^\text{recorded} - \rho_{\text{sync}}^\text{recorded}} \]
Estimating uncertainty

To estimate the uncertainty of TCP in our analysis, 2000 Monte Carlo samples were drawn randomly from beta distributions representing the mean and 95% binomial confidence intervals (CIs) of the 5-year EFS values of each study included in the model. For each sample, a unique dose-response curve was derived and used to estimate the TCP. It should be noted that this does not factor in potential systematic errors like e.g. the true shape of the tumor control dose-response curve being different from what is assumed in our model.

Results

Figure 3 shows the resulting tumor control dose-response models. The spread in sampled data points illustrates that tumor control estimates for CSI doses ranging from about 23 to 33 Gy are fairly robust while for lower doses there is considerable uncertainty. The boost dose in the studies with 18 Gy CSI varied between 48 to 54 Gy.

Figure 3. Dose-response curves for estimating tumor control in the posterior fossa boost site (left) and the elective craniospinal site (right). For graphical clarity, only 200 randomly sampled curves are shown. There is large variation in the data representing the studies with 18 Gy craniospinal dose (and corresponding boost dose) which reflects the sparseness of outcome data at these doses. Dashed lines show the 95% confidence limits of the dose-response curves based on 2000 Monte Carlo samples.
The parameters corresponding to the median estimate of the sampled dose-response models fitted to the data in Table 1 were, \( \gamma_{50, \text{elective}} = \gamma_{50, \text{boost}} = 0.36 \) (fixed), \( D_{50, \text{elective}} = 62.5 \text{ Gy} \), \( D_{50, \text{boost}} = 13.1 \text{ Gy} \), \( P_{0, \text{elective}} = 0.857 \) and \( P_{0, \text{boost}} = 0.00 \). For CSI doses of zero, 18, 24 and 36 Gy, the estimated values on 5-year FFP (95% CI) based on 2000 Monte Carlo samples of the dose-response parameters were 76\% (0-81\%), 78\% (73-81\%), 79\% (76-82\%) and 80\% (77-84\%), respectively while keeping the PF dose constant at 54 Gy. Estimates of the “cured fraction” of patients \( (P_c) \) were calculated from Eq. 4, cf. Figure 4. The effect on FFP from changing the CSI dose is illustrated in Figure 5.

**Figure 4.** The “cured” fraction \( (P_c) \) estimated as in Eq. 4) i.e. the proportion of patients estimated to be persistently free from disease progression, as function of dose to the boosted volume (left) while maintaining 24 Gy to the craniospinal elective volume and as function of dose to the elective volume (right) while maintaining 54 Gy to the posterior fossa. The solid line represents the point estimates and the shaded areas encompassed by the dashed lines give the 95\% confidence limits based on 2000 Monte Carlo samples.
Figure 5. Estimated freedom from progression (FFP) for three scenarios, zero Gy craniospinal irradiation (CSI), 24 Gy CSI or 36 Gy CSI, in all scenarios keeping the dose to the boost site constant at 54 Gy. Model parameters used were $\alpha = 1.99$ and $\lambda = 2.39 \, \text{y}^{-1}$, while $P_-$ varies with dose. Vertical bars represent the 95% confidence limits for FFP at 5 years after diagnosis (displaced slightly in the graph for visualization), based on 2000 Monte Carlo samples.

Discussion

A framework for modeling tumor control with multiple modes of failure was developed for standard-risk MB. A mixture model was implemented to convert TCP estimates to FFP up to 10 years after diagnosis. Recent attempts to alter the treatment of standard-risk MB, have either been with the aim of lowering toxicity or improving efficacy (1, 2, 4, 7, 8). Modeling the 10-year risk of relapse provides the possibility to generate data driven hypotheses by simulating altered treatment regimens e.g. using modern radiotherapy planning software. One immediate application is to provide data-driven power calculations for MB trials.
Application to a post-hoc power calculation for the standard- vs. hyperfractionated treatment compared in the international multi-center PNET-4 trial yields an estimated 5-year FFP (95% CI) of 78.3% (76.5-82.5%) in the standard arm and 81.2% (77.6-89.3%) in the hyperfractionated arm. The sample size required to show this difference at a 5% confidence level (risk of type I error) and 90% statistical power (10% risk of type II error) is 3,287 patients in each arm. The PNET-4 study enrolled 169 patients in each treatment arm (1). Based on our model, the statistical power for detecting a difference would be about 10%, consistent with the trial's negative result. Thus, in a post-hoc setting, our FFP model does predict that the outcome in the two arms in PNET-4 was not sufficient to show a statistically significant difference given the sample size. Since the result from PNET-4 was included in our model this does not constitute a model validation. It does, however, show the importance of very large multi-center efforts and illustrates the relevance of TCP models for clinical trial design.

It is important to consider the potential difference in failure pattern between different molecular subgroups. A recent study showed the effectiveness of including molecular biomarkers, in addition to clinical variables, to predict the risk of relapse (11). The effectiveness of radiotherapy may differ between subgroups and although our model estimated a shallow dose-response for the elective CSI in standard-risk patients, the dose-response curves are likely steeper when separately analyzing subgroups (cf. prostate cancer) (19). Unfortunately, such outcome data are not yet available for molecular stratification. Access to detailed pattern of failure data for different MB subgroups and varying treatment doses would enable synthesizing dose-response models stratified on molecular markers (9). Hopefully, secondary analysis of the recent randomized trials could provide insight into this.

Interpreting the results of this modeling study should be done with caution. There are considerable gaps in the knowledge of the efficacy of low craniospinal doses, illustrated by the very large uncertainty in outcome estimates at low doses. Also, the possibility that the functional form of the models might be different from the true underlying dose-response relationship cannot be factored into the estimated uncertainty.

Assuming a fixed \( \gamma_{50} \) of 0.36 based on the study by Thomas et al. (5) means assuming that the steepness of the dose-response curve based on their comparison of 23.4 and 36 Gy CSI without
chemotherapy is also valid for studies including chemotherapy. Applying this fixed $g_{50}$ value in our model may introduce some further uncertainty into the dose-response model which cannot be considered in the estimated statistical uncertainty. However, this is not expected to have a substantial impact on the wide confidence intervals currently considered, mainly stemming from the uncertainty in the published 5-year EFS data. Our results indicate that a very low (or even zero) dose-response in the elective target volume could be a possibility. This is not the case when treatment is given without chemotherapy. In a previous MB report there was a clear improvement in outcome when moving from posterior fossa irradiation only, to including also the spinal subdural space with or without further inclusion of the supratentorial subdural space (20). This supports the hypothesis that the shallow elective volume dose-response in standard-risk MB could be attributed the effectiveness of chemotherapy in handling sub-clinical disease.

Estimating tumor control with multiple modes of failure requires dealing with failures that are recorded as synchronous in multiple target sites. Here, it was assumed that failure in the two sub-sites were independent. It is possible that this assumption does not hold, e.g. if patients failing in the primary site have a more malignant phenotype than patients not failing after radiotherapy. Appendix A (see supplementary material) presents a sensitivity analysis which showed that relaxing the assumption of independence had only a small effect on our model predictions. Given detailed pattern of failure studies from individual patient data, including molecular biomarkers, competing risk methods could potentially help in understanding the relationship between failures in the different sites.

Ideally, a dose-response and time-to-progression model would be derived from a single randomized trial cohort, with patients randomly allocated to different dose prescriptions. In the absence of such data we synthesized the knowledge from existing studies using a meta-analysis approach. Potential pitfalls could be that patient-related characteristics or non-radiotherapy treatment factors change over time. The possibility that the risk of failure might be non-uniformly distributed throughout the different target volumes was outside the scope of this study and cannot be tested with the currently available reports on patterns of failure. On the other hand, the methods developed here may be extended to generate a model of TCP based on more detailed pattern of failure analyses in the future. We acknowledge the uncertainties in the presented model, and the lack of clinical data in the low dose range.


Appendix A

When evaluating the potential clinical benefit from a strategy of redistributing dose between the elective and boost volume, it is important to consider whether a recurrence in the elective volume that is detected at the same time as a recurrence in the boost volume is truly a synchronous recurrence or whether it is secondary to a local failure in the boost volume that has been subclinical for a certain period. A sensitivity analysis is performed where the assumption of statistical independence between the probabilities of recurrence in the elective and the boost volume is relaxed. This is achieved through the introduction of a parameter $\eta$ where a value of 1 corresponds to statistical independence and a value lower than 1 represents a higher number of truly synchronous recurrences.

\[ \rho_{\text{true}}^{\text{elective}} \cdot \rho_{\text{true}}^{\text{boost}} = \eta \cdot \rho_{\text{true}}^{\text{sync}}, \text{ where } \eta \leq 1 \]

\[ \rho_{\text{true}}^{\text{elective}} = \rho_{\text{recorded}}^{\text{elective}} + \rho_{\text{true}}^{\text{sync}} \]

As we assume that the recorded synchronous failures are either truly synchronous or failures arising from the boost:

\[ \rho_{\text{true}}^{\text{elective}} = \rho_{\text{recorded}}^{\text{elective}} + \rho_{\text{true}}^{\text{sync}} \]

Since

\[ \rho_{\text{true}}^{\text{elective}} \cdot \rho_{\text{true}}^{\text{boost}} = \eta \cdot \rho_{\text{true}}^{\text{sync}} \]

this gives:

\[ \left( \rho_{\text{elective}} + \rho_{\text{true}}^{\text{sync}} \right) \left( \rho_{\text{recorded}}^{\text{boost}} + \rho_{\text{true}}^{\text{sync}} \right) = \eta \cdot \rho_{\text{true}}^{\text{sync}} \]

\[ \rho_{\text{elective}} \left( \rho_{\text{recorded}}^{\text{boost}} + \rho_{\text{true}}^{\text{sync}} \right) = \rho_{\text{true}}^{\text{sync}} \left( \eta - \rho_{\text{true}}^{\text{recorded}} - \rho_{\text{true}}^{\text{sync}} \right) \]

\[ \rho_{\text{true}}^{\text{sync}} = \frac{\rho_{\text{true}}^{\text{elective}} - \rho_{\text{true}}^{\text{recorded}}}{\eta - \rho_{\text{true}}^{\text{recorded}} - \rho_{\text{true}}^{\text{sync}}} \]

We find the lowest value of $\eta$ that satisfies the condition $\rho_{\text{true}}^{\text{elective}} \leq \rho_{\text{true}}^{\text{recorded}}$ when analyzing the pattern of failure data from the two largest studies included in our dose-response model (Packer 2006 and PNET-4 2012). This condition gives the lowest $\eta$ consistent with the data since an even lower value would yield a number of truly synchronous failures that is larger than the number of recorded synchronous failures which would be inconsistent with the fact that patients are only examined for potential disease recurrence at certain time intervals and not continuously. In other words, this corresponds to the model which assigns the largest possible number of failures to the elective volume. The number of failures in the primary site remains unchanged. This value was found to be $\eta = 0.72$ and inputting this into the model yields a $D_{50, \text{elective}} = 64.6$ Gy and $P_0, \text{elective} = 0.86$ not affecting the estimated cured fraction for the elective volume noticeably.
STUDY VI
Risk-based radiation therapy optimization through simultaneous common scale comparison of the life years lost attributable to tumor control and late complication risk

N. Patrik Brodin¹,², Ivan R. Vogelius¹, Thomas Björk-Eriksson³, Per Munck af Rosenschöld¹,², Maja V. Maraldo⁴, Marianne C. Aznar¹,², Lena Specht⁴, Søren M. Bentzen⁵

¹Radiation Medicine Research Center, Department of Radiation Oncology, Rigshospitalet, Copenhagen, Denmark
²Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
³Department of Oncology, Skåne University Hospital and Lund University, Lund, Sweden
⁴Department of Radiation Oncology, Rigshospitalet, Copenhagen, Denmark
⁵Department of Human Oncology, University of Wisconsin Medical School, Madison, WI, USA

Funding

This work was supported by a grant from the Danish Child Cancer Foundation
Abstract

Background: A mathematical framework is presented for simultaneous quantification and evaluation of the trade-off between tumor control (TC) and late complications for risk-based decision-making in radiation therapy (RT). We estimate the life years lost (LYL) attributable to tumor recurrence, late cardiac toxicity and secondary cancers to compare the expected clinical effect of re-distribution of radiation dose on a common scale. An example of this risk-based treatment evaluation is presented for standard-risk pediatric medulloblastoma (MB).

Methods: A bottom-line LYL estimate was derived for standard risk MB, based on the LYL attributable to the risk of radiation-induced late complications and the LYL due to failure to control the primary disease. We compared the standard approach to RT for MB using 3D conformal photon therapy with 1) proton therapy and 2) risk-adaptive photon treatment strategies lowering the prescribed dose to part of the craniospinal (CS) target volume situated close to critical risk organs. This was done through treatment planning simulation of 10 pediatric MB patients.

Results: Late toxicity is important, with 0.75 years (95% CI: 0.60-7.2 years) LYL for the current standard uniform 24 Gy CS dose. However, the LYL attributable to recurrence of the primary disease dominates the total number of LYL accounting for 14.2 years lost (95% CI: 13.4-16.6 years). Compared to the standard treatment, a risk-adapted strategy of prescribing 12 Gy to the part of the target encompassing the 1st to the 10th thoracic vertebrae, while treating the remaining CS volume to 36 Gy resulted in a LYL reduction of 0.90 years (95% CI: -0.18 - 2.41 years). This demonstrates that lowering the target dose in the vicinity of critical risk organs can result in a net reduction in estimated LYL, if a small reduction in TC is outweighed by the reduced risk of late complications.

Conclusions: Risk of recurrence of the primary disease dominated the total LYL of pediatric MB patients. It was possible to optimize a radiation prescription strategy simultaneously adapted to the risk of normal tissue toxicity and risk of recurrence - an all-cause mortality dose painting approach. The risk-adapted techniques compared favorably to the standard, but in the current context, the absolute gain is small compared to the underlying uncertainty in the model.
**Introduction**

A key challenge for clinicians in radiation therapy (RT) is to prioritize between the chance of cure and the risk of severe treatment complications. This trade-off varies between patients and can also depend on available treatment modalities. Yet, few quantitative and testable decision making tools have been proposed to aid clinicians in the process of prioritizing risks across different endpoints. As pediatric patients are especially sensitive to treatment-induced complications, there is a great need for such tools as most children will now be cured from their disease and become long-term survivors.\(^1\) Hence, the focus of radiotherapy research has recently shifted towards reducing treatment-related toxicity while maintaining a high rate of disease control for pediatric patients. Quantifying the trade-off between tumor control and the risk of late complications is far from trivial since it involves comparing measures of primary disease survival with estimated risks of e.g. developing secondary cancer or cardiac disease up to several decades later. Here, we propose using a life years lost (LYL) measure for common scale comparison of the risk of different late complications with that of failing to control the primary disease. This measure naturally accounts for the difference in time-to-event, severity of the given event and the presence of competing risks between the different endpoints.

For homogenous targets, a highly conformal and uniform radiation dose distribution is optimal in terms of tumor control.\(^2\) An overall better treatment could, however, be associated with a slight reduction in tumor control probability (TCP) if, in return, it would considerably reduce the incidence of severe late toxicity. Using the LYL measure we set up a mathematical framework to investigate if limiting the prescribed radiation dose to a part of the target volume situated close to critical risk organs and thus reduce the risk of late complications could outweigh the potential increase in relapse risk. Prescription of non-uniform doses to the target based on heterogeneity in tumor radio-responsiveness has already been the subject of numerous studies. This is, to our knowledge, the first attempt at defining an optimal, non-uniform dose based on the risk of normal tissue complications in the presence of an (assumed) homogeneous risk of local tumor recurrence within the radiotherapy target volume.

Here, we present a proof of principle study for such a decision-making tool tested on a group of pediatric medulloblastoma (MB) patients for a set of risk-adapted treatment strategies. Specifically, we aim to evaluate if the LYL can be reduced with a non-uniform dose prescription and estimating the uncertainty of the total LYL estimates.
Methods and materials

Concept methodology - Estimating the LYL attributable to the primary disease and late complications

The methodology for estimating the LYL attributable to late complications for patients undergoing radiotherapy has been described in detail elsewhere\(^3\), although some of the dose-response models have been updated (dose-response models used in this analysis are given in Appendix A). Late complications considered to attribute to the LYL in this analysis were secondary breast, lung, stomach and thyroid cancer as well as heart failure and myocardial infarction.

A tumor control dose-response model is used to estimate the 5-year OS as a function of radiation dose. This must be combined with a survival function after disease relapse/progression in order to estimate disease-specific survival (DSS) after treatment as a function of dose.

In the following, we derive an expression for DSS as a function of time. The probability of relapsing at an attained age \(a\) is the product of the probability of being failure free up to time \(a\) (here denoted as freedom from progression (FFP(\(a\))) and the dose-dependent hazard of relapse at time \(a\), \(h_{\text{relapse}}(a)\);

\[
P_{\text{relapse}}(a) = \text{FFP}(a) \cdot h_{\text{relapse}}(a)
\]  

(1)

The risk of dying within a narrow time interval from \(t \rightarrow t + \Delta t\) from a relapse occurring at time \(a\) is defined as

\[
S_{\text{relapse}}(t - a) - S_{\text{relapse}}(t + \Delta t - a)
\]

(2)

where \(S_{\text{relapse}}\) is the survival probability after relapse. We parameterize this survival function as an exponential, \(S_{\text{relapse}}(t) = \exp(-\lambda_{S} \cdot t)\) fitted to data on the survival after relapse. Here \(\lambda_{S}\) is the constant hazard of dying after a relapse and \(t\) is the time passed since relapse. The disease-specific risk of dying in the time interval \(t \rightarrow t + \Delta t\) after a recurrence occurring at time \(a\) is then

\[
\text{FFP}(a) \cdot h_{\text{relapse}}(a) \cdot \left[ S_{\text{relapse}}(t - a) - S_{\text{relapse}}(t + \Delta t - a) \right]
\]

(3)

We then integrate over all possible relapse times, \(a\), ranging from time of treatment (age at exposure, \(e\)) to time of observation, \(t\), to find the total risk of dying from a relapse in the narrow time interval \(t \rightarrow t + \Delta t\).
\[ DSS(t) - DSS(t + \Delta t) = \int_{a=e}^{t} \text{FFP}(a) \cdot h_{\text{relapse}}(a) \cdot \left[ S_{\text{relapse}}(t-a) - S_{\text{relapse}}(t+\Delta t-a) \right] da \]  (4)

We correct for the (very small) non-cancer related competing risks by multiplying the calculated DSS with the age- and sex-matched survival in the US general population to yield the OS shown in Figure 1.

**Figure 1.** Impact of the estimated tumor control probability (TCP) on the survival curve. The TCP affects the LYL of late complications because these are competing risks - a poor long-term survival means less risk of experiencing a late event. This example shows the difference in estimated overall survival for males and females for a TCP of either 80% or 60%.

We chose to calculate the disease-specific risk of dying up to 10 years after treatment, as most relapse-related deaths will have occurred by then, cf. Figure 2. That means that until \( t = e + 10 \) years we use Eq. 4 to calculate the DSS. More than 10 years after treatment, the OS is described by a survival function with long-term follow-up of a cohort of patients relevant to the disease in question, preferably including death due to late recurrence and other potential causes of death, normalized to fit the TCP based OS function at 10 years.
Figure 2. The estimated risk of dying from failure to control the primary disease using our presented mathematical framework is shown to increase until about 10 years after diagnosis. The graphs compare four different levels of tumor control probability (TCP), which corresponds to the 5-year freedom from progression.

The LYL attributable to the primary disease (denoted LYL_{relapse}) is then estimated for the first 10 years after diagnosis. The LE after having relapsed (LE_{relapse}) is derived from the survival curve after relapse, S_{relapse}(t). The LYL from a relapse occurring at attained age \(a\) is the probability of having a relapse, multiplied by the difference in LE for someone with a relapse of the disease in question compared to that of an age- and sex-matched person in the general population (LE_{gen.pop}):

\[
LYL_{relapse}(a) = P_{relapse}(a) \cdot (LE_{gen.pop}(a) - LE_{relapse}(a))
\]

(5)

where \(LE_{relapse} = \int_0^\infty t \cdot \exp(-\lambda \cdot t) dt\) and the \(LE_{gen.pop}\) is taken from life tables of the US general population. The total LYL attributable to the primary disease is then obtained by integrating over all ages, \(a\), up to 10 years after treatment.

\[
LYL_{relapse} = \int_\epsilon^{10+\epsilon} P_{relapse}(a) \cdot \left[ LE_{gen.pop}(a) - \int_0^\infty t \cdot \exp(-\lambda \cdot t) dt \right] da
\]

(6)
Application to standard-risk medulloblastoma

In order to apply the above methodology for pediatric MB we thus need a time-dependent tumor control dose-response model, a survival function after relapse, dose-response models for radiation-induced late complications and long-term survival data for MB patients.

A tumor control dose-response model estimating the 5-year FPP for standard-risk MB patients which considers multiple prescription levels and modes of failure was derived and explained in detail elsewhere (Manuscript: Modeling freedom from progression for standard-risk medulloblastoma: A mathematical tumor control model with multiple modes of failure (Study V)). In order to generalize this TCP model to inhomogeneous dose distributions, we here assumed that:

- The risk of relapse is uniformly distributed throughout each of the two target volumes, (primary target of radiation (boost volume) and the elective craniospinal volume), respectively
- The relapse risk in all volume elements (i.e. voxels) are independent of each other

The formulation suggested by Kim and Tomé for calculating an overall TCP based on estimates of several sub-regions was applied:

$$TCP(D) = \prod_{j=1}^{R} TCP_{j} \left( \left( D_{j}^{i}, v_{j}^{i} \right) \right) = \prod_{j=1}^{R} \prod_{i=1}^{k_{R}} \left( TCP_{j}^{i} \right)^{v_{j}^{i}} \quad (7)$$

where $D_{j}^{i}$ is the dose to the $i$:th voxel of the $j$:th target volume and $v_{j}^{i}$ is the fractional volume of that voxel. $R$ is the number of sub-regions of the target and $k_{R}$ is the number of voxels within a sub-region.

The assumption of independence is used to define the total TCP as the product of the control probability in all voxels and sub-regions. The dose-response relations for the different sub-regions were derived as logistic functions with the logarithm of dose as a covariate, as presented in Bentzen and Tucker.6

$$TCP_{j}^{i} = \frac{1 - P_{0}}{1 + \left( \frac{D_{50}}{D} \right)^{4\gamma_{50}}} + P_{0} \quad (8)$$

where $D_{j}^{i}$ represents the dose received by the voxels in $v_{j}^{i}$ which is the $i$:th fractional volume in the $j$:th sub-region of the target. In this case there are two sub-regions ($R = 2$), the elective volume and the boost volume, for which their individual dose-volume histograms are made up of $k_{R}$ volume elements. $\gamma_{50}^{i}$ and
$D_{50}^j$ describe the logistic dose-response curve for each sub-region, $P_0$ is tumor control with no radiation. The TCP estimates calculated from Eq. 6 corresponds to the FFP up to 5 years after treatment.

The survival probability for relapsing MB patients ($S_{\text{relapse}}$) treated with high-dose chemotherapy or standard salvage therapy was derived from a previously published study with an exponential model was fitted to data and the long-term survival for MB patients was taken from the survival curve presented for survivors of CNS malignancies in the CCSS.

**Patient material and treatment planning strategies**

Treatment plans with different prescription strategies were simulated on computed tomography scans of 10 pediatric MB patients aged 4-15 years treated at Rigshospitalet, Copenhagen in 2007-2009. In MB treatment, the risk of treatment-related side-effects has been reduced somewhat through the introduction and wide acceptance of a lowered craniospinal (CS) treatment dose to about 24 Gy for standard-risk patients, compared to the previous prescription of 36 Gy. Good disease control rates are achieved for the standard-risk group with this lower CS dose with 5-year event-free survival (EFS) ranging from about 75-85%.9,10 To investigate the possibility of risk-based decision-making, treatment plans were generated using 3D conformal radiation therapy (3D CRT) where the following treatment strategies were explored:

1. Homogeneous dose distribution of 24 Gy using 3D CRT (current standard)
2. Lowering the prescribed dose to the part of the CS target encompassing the 1st to 10th thoracic vertebra (Th1-Th10) using 3D CRT; the rationale being that this region would include the most critical structures (heart, lungs, breasts)
3. Lowering the prescribed dose to the entire spinal canal using 3D CRT, which would include all risk organs related to the late complications considered in this analysis
4. Homogeneous dose distribution using proton therapy; to compare the risk-adaptive prescription approach with the healthy tissue sparing achievable with proton therapy, without reducing the dose to the target volume.

For all strategies, the cranial target was planned with two lateral opposing fields. For strategy 1) and 3) a single spinal field from the posterior direction was planned and for strategy 2) we used two separate spinal fields. Four fields from the latero-posterior direction (two opposed and two oblique) were used to plan the posterior fossa (PF) boost, except for proton therapy where only one posterior field was used. Details for the planning of the spot-scanned proton therapy plans are described elsewhere.11 Secondary neutron doses were estimated for the proton plans according to the same method as in the previous report.
although here we used the recommended values of the relative biologic effectiveness of neutrons from ICRP publication 92.

Treatment plans were simulated with the prescribed dose varying between 0 to 36 Gy in 3 Gy increments, resulting in 13 different plans for each patient. Similarly to the Th1-Th10 or spinal canal, we also varied the dose to the remaining part of the CS target volume in the same way. This was done to investigate whether a potential drop in TCP from lowering the dose to part of the target could be compensated for by increasing the dose to the remaining target volume. The dose to the PF boost volume was kept constant at 54 Gy for all scenarios. All in all, 13x13 treatment plans were simulated for each of the 10 patients and for each of the two prescription strategies and analyzed according to estimated TCP and the number of LYL, attributable to the primary disease and late complications.

Statistical analysis

The major source of uncertainty in the LYL estimates stems from the dose-response parameters, both for the TCP model and for the late complications. To assess the uncertainty in the final LYL estimates, a Monte Carlo sampling method was used, sampling over the uncertainty in all of the dose-response parameters. For the TCP model, 2000 samples were drawn randomly from beta distributions representing the mean and 95% binomial confidence intervals (CIs) of the 5-year EFS values of each study included in the model. (Manuscript: Modeling freedom from progression for standard-risk medulloblastoma: A mathematical tumor control model with multiple modes of failure (Study V)) For each sample, a unique dose-response curve was derived and used to estimate the TCP. Similarly, 2000 samples were drawn from log-normal distributions representing the mean and 95% CIs of the dose-response parameters for the late complications (given in Appendix A). This provides us with statistical uncertainty estimates based on the input parameter uncertainties. What is not considered in this uncertainty however, are potential systematic errors. Examples of such errors could be a breakdown of the assumptions for the TCP model not being valid or the true functional form of the risk-estimation curves being different from the ones applied in this study.

For each randomly drawn set of model parameters, the LYL attributable to the disease and the late complications were calculated for the whole patient population and all the treatment prescriptions studied. The final estimates of LYL and 95% CI for each patient were taken as the median and 2.5 – 97.5 percentile of the resulting LYL distributions. Similarly, the differences in LYL between treatment strategies was found by calculating the LYL differences for each randomly drawn sample and extracting
the difference estimate with 95% CI for each patient in the same way, i.e. a paired comparison. To avoid potentially underestimating the variance due to the small number of patients, a bootstrapping procedure was applied where the median and 95% CI was derived for 200,000 samples of the ten patients randomly drawn with replacement. For each sample, the estimates were weighted by the inverse variance between patients and the final difference in LYL with 95% CI was taken as the median and 2.5 - 97.5 percentile of all randomly drawn bootstrap samples.

**Results**

Figure 3 compares the estimated number of LYL between treating the whole CS volume to a uniform high dose of 36 Gy and a risk-adapted strategy of lowering the dose to the Th1-Th10 to 12 Gy. It clearly shows that the LYL attributable to the primary disease has a major impact on the resulting LE. However, reducing the dose to the Th1-Th10 target to 12 Gy lowers the estimated number of total LYL by substantially decreasing the risk of late complications while the accompanying drop in TCP is estimated to be very small.

For CS doses of zero, 18, 24 and 36 Gy the average estimated TCP values, and 95% CIs based on 2000 Monte Carlo samples of the dose-response parameters, representing 5-year FFP were 76% (0-81%), 78% (73-81%), 79% (76-82%) and 80% (77-84%) respectively, while maintaining a 54 Gy PF dose. The estimated 5-year FFP with zero Gy to the Th1-Th10 target or to the entire spinal canal while maintaining 24 Gy to the rest of the CS volume was 78% (63-81%) or 78% (8-81%), respectively. This is a result of the median dose-response model for the CS elective volume being quite shallow, although with very large uncertainty given by the wide CIs.
Figure 3. Number of life years lost attributable to various late complications and failure to control the primary disease for two different prescription strategies. Even though the primary disease dominates the life years lost there is an estimated gain if lowering the dose to the Th1-Th10 part of the craniospinal target. In the right hand panel, the anatomical position of the heart, lungs, thyroid, mammary glands and stomach are shown in relation to the craniospinal target (in cyan) for one of the patients.

Reducing the dose to the whole CS target yields a net increase in the estimated number of LYL, whereas reducing the dose to the Th1-Th10 sub-target only causes a small drop in TCP that is compensated for by a reduced risk of late complications, cf. Figure 4. It should be noted that the 95% CIs are quite large, especially for low doses to the whole CS volume which stems from considerable uncertainty in the TCP dose-response model at such low doses.
Figure 4. Estimated total number of life years lost if altering the prescription dose to (a) the whole craniospinal target volume or (b) Th1-Th10. Here (b) corresponds to treatment strategy 2 in the text. Confidence limits are based on Monte Carlo sampling of the dose-response parameters for late complications and tumor control.

Figure 5. Separation of the life years lost (LYL) attributable to late complications and failure to control the primary disease. Here, the dose to the spinal canal is altered (treatment strategy 3) while a dose of 24 Gy is maintained to the remaining CS volume. Confidence limits are based on Monte Carlo sampling of the dose-response parameters for late complications and tumor control. Note that it is the failure to control the primary tumor that dominates the total LYL for these patients.
The effect of altered dose prescriptions on the combined uncertainty in LYL attributable to the primary disease and to late complications can be seen in Figure 4 and Figure 5. We can clearly see that the total LYL is dominated by failure to control the primary disease (Figure 3 and 5).

The prescription dose yielding the minimum number of LYL was found at giving zero Gy to the Th1-Th10 while treating the remaining CS volume to the full dose of 36 Gy. This is a result of the shallow TCP dose-response curve where the increase in LYL attributable to lower tumor control is outweighed by the decrease in LYL from a reduced risk of late complications. If prescribing zero Gy to either the Th1-Th10 or the spinal canal the LYL attributable to disease control failure were 13.5 years (95% CI: 11.8-19.5 years) and 13.7 years (95% CI: 12.2-32.8 years) respectively. The high upper limits of these CIs reflect the very large uncertainty in TCP estimates at low doses.

Figure 6 shows the estimated difference with 95% CIs in LYL between the standard uniform 24 Gy treatment and 5 specific risk-adapted strategies.
Figure 6. Differences in total life years lost (LYL) between the standard uniform 24 Gy treatment and risk-adapted strategies. For all risk-adaptation strategies, 36 Gy is prescribed to the remaining craniospinal volume. Accompanying illustrations of the different treatment strategies are shown in dose color-wash. A positive LYL difference means fewer LYL with the risk-adapted approach. The vertical bars show the 95% confidence intervals of the LYL differences.

Here it shows that prescribing 12 and 18 Gy to the Th1-Th10 and spinal canal respectively results in slightly less life years spared when compared to prescribing zero Gy to these areas. However, the 12-18 Gy prescriptions have considerably less uncertainty, especially in the unfavorable end of the CI. A homogeneous dose of 36 Gy with protons yields an estimated LYL from the primary disease of 13.3 years (95% CI: 10.8-14.7 years) and 0.12 years (95% CI: 0.04-0.57 years) from late complications. In a paired statistical comparison, there was significantly more life years spared with proton therapy compared to the risk-adapted photon strategies. The mean differences were 0.50 years (95% CI: 0.25-2.60 years) between
the 12 Gy to the Th1-Th10 strategy and proton therapy and 0.68 years (95% CI: 0.36-2.57 years) between the 18 Gy to the spinal canal strategy and proton therapy.

**Discussion**

A method for comparing estimates of TCP and risks of late complications on a common scale was developed and applied to assess risk-adaptive prescription strategies for pediatric MB patients. A common scale measure of complication risk and TCP has two advantages; 1) it allows direct quantification of the therapeutic ratio between different treatment or prescription strategies and 2) it provides an understanding of the relative importance of each endpoint. It has long been recognized that late morbidity and mortality is a major issue for childhood cancer patients. Our data reflects this fact, but also emphasizes that lack of tumor control is by far the dominant cause of LYL and, consequently, strategies to decrease the risk of late complications should be carefully weighed against the potential loss in tumor control.

We performed a simple optimization of the dose prescription to show that a homogeneous dose to the target may not be the optimal choice, even for a homogeneous tumor recurrence risk per unit volume, when also considering the effect on healthy tissues. This is, to our knowledge, the first attempt at presenting a non-uniform dose prescription strategy based on risk of toxicity rather than intra-tumor heterogeneity. This adds a dimension to the famous demonstration by Brahme et al. that a homogeneous tumor is best irradiated with a homogeneous dose. The prescription strategies explored here focused on lowering the dose to the thoracic region. Considering also non-lethal endpoints such as neurocognitive decline from the cranial irradiation would be an interesting extension, but is beyond the scope of this study. Admittedly, the uncertainties in our model estimates are considerable, so the present study should be seen as a proof-of-principle study rather than an attempt to derive an alternative prescription strategy for clinical use.

The uncertainty in the estimated TCP is lowest around CS doses of 24-36 Gy since this is where most clinical experience is concentrated. Consequently, the uncertainty in LYL attributable to failing to control the primary disease is also lowest around this dose. There have been attempts to further lower the CS dose to 18 Gy where the conclusion was that the rate of relapse was unacceptably high at this dose, although the patient numbers were very limited. The results of the ongoing ACNS 0331 trial which randomizes MB patients between 23.4 Gy and 18 Gy CS irradiation should provide further insight into this. The uncertainty in LYL attributable to late complications, however, is roughly proportional to dose, which is a result of all but one risk model adhering to a linear dose-effect relationship (cf. Appendix A).
When interpreting the results of this proof of concept analysis it is important to consider the uncertainties and assumptions involved, especially in the TCP model. The assumptions need to be validated, preferably in a large cohort of MB patients with detailed information on treatment and patterns of failure. Whether the risk of relapse is in fact uniformly distributed within the craniospinal target volume cannot, currently, be deduced from reported pattern of failure data.\textsuperscript{9, 10} The pattern of failure also appears to vary considerably between different reports on standard-risk medulloblastoma patients, cf. Table 1 in Brodin et al. (Manuscript: \textit{Modeling freedom from progression for standard-risk medulloblastoma: A mathematical tumor control model with multiple modes of failure (Study V)}).

From mortality data from the CCSS cohort it was estimated that, with a mean age of 8 years at diagnosis, MB patients that are 5-year survivors have an estimated LE of 47 years.\textsuperscript{15} From the parametric TCP-based survival functions derived in this analysis the estimated LE of an 8 year old MB patient (average between males and females), conditional upon being a 5-year survivor, is 49 years. This shows that fitting the post 10-year survival curve to the last point of the parametric survival function gives long-term survival estimates that are in agreement with previously published data.

Since MB has recently been characterized as comprising of four distinct molecular subgroups\textsuperscript{16}, it is likely that different strategies for altering the radiotherapy dose should be applied to the different subgroups. As a step towards further personalizing treatment, the analysis performed in this study could be repeated for each individual subgroup, conditional on detailed outcome data from recent MB trials, stratified on molecular biomarkers, becoming available.

Implementing this LYL model directly into treatment planning software would allow for direct risk-based optimization with the aim of minimizing the total number of LYL. While the potential benefit in terms of LYL uncovered in this study may not be of sufficient magnitude to warrant such a strategy, the fundamental principle of simultaneous optimization of TCP and normal tissue endpoints on a common scale is very attractive. Among patients who could potentially have more benefit from such a treatment optimization approach are those with Hodgkin’s lymphoma, because of their excellent prognosis, good salvage options, and considerable risk of severe late complications.\textsuperscript{17}

For the parts of the craniospinal volume that are not located close to the critical healthy tissues considered in this study (e.g. the brain), the optimal treatment, according to the model, would be to prescribe as much dose as possible to these areas. Although far from all potential radiation-related treatment complications can be modeled, they should be considered when critically evaluating the potential of such a direct risk-based optimization approach. A limitation of the current analysis is that non-lethal endpoints such as
cognitive decline or endocrine complications are not considered here and also including such complications would more closely resemble the clinical situation.

In conclusion, we have shown how endpoints related to controlling the primary disease and late complications can be evaluated on the same scale. We have also shown that this provides an opportunity for simultaneous optimization of tumor control and normal tissue toxicity, where the mathematical framework provided here can be applied to other cancer sites given the required input data are available. This results in an optimal, heterogeneous dose distribution, corresponding to an all-cause mortality driven dose painting.
Reference List


