Potential benefits of intensity-modulated proton therapy in head and neck cancer

van de Water, Tara Arpana

Chapter 8

General discussion and conclusions

General discussion and conclusions

The aim of this thesis was (1) to investigate if irradiation with scanned intensity-modulated proton therapy (IMPT) in head and neck cancer (HNC) would result in significant dose reductions to the most relevant organs at risk (OARs) and (2) to estimate whether these dose reductions would theoretically translate into a reduction of radiation-induced side effects in relation to the current standard, i.e. intensity-modulated radiotherapy with photons (IMRT).

In this project, we focussed on the two most important side effects of curatively intended radiotherapy and chemoradiation in the head and neck area, i.e. xerostomia and swallowing dysfunction. Xerostomia is the most frequently reported grade 2 or higher side effect of radiotherapy in the head and neck region, while swallowing dysfunction is the most frequently reported grade 3 side effect [105]. Moreover, these two side effects have a significant adverse impact on health-related quality of life as reported by patients [47,105,130,146,186]. As such, these two side effects are clinically relevant for patients.

Methodology

The department of Radiation Oncology of the University Medical Center Groningen (UMCG) developed a 3-step methodology aiming at defining cohorts of patients who may benefit from protons by using the obtained dose distribution results of the individual *in silico* planning comparative (ISPC) studies as input in normal tissue complication probability (NTCP) models.

More specifically, this methodology consists of: (1) the development and validation of NTCP models; (2) *in silico* planning comparative studies, and (3) estimation of the clinical benefit of protons in reducing radiation-induced side effects by translating the dose distribution results of the individual ISPC studies into estimated clinical benefits by using NTCP models. In the following paragraphs, these three steps are described in more detail.

STEP 1: The development of normal tissue complication probability models

The basic principle in the development of new radiation delivery techniques is the relationship between dose distributions in OARs and the probability of radiation-induced side effects (i.e. the NTCP). In general, the estimated risk on a given side effect, i.e. the NTCP value, will increase with increasing dose and increasing volume that receives a certain dose (see Figure 1 chapter 3). For the purpose of this methodology, a prerequisite is that at least one dose distribution parameter significantly correlates with the risk on a given side effect. If this is the case, this parameter can be used for optimisation of the radiation technique (i.e. optimising the treatment plan) in STEP 2.

STEP 2: In silico planning comparative (ISPC) studies

With respect to reduction of side effects, the potential benefit of proton radiotherapy is mainly based on the assumption that this new technique provides an improved dose distribution, i.e. an at least equivalent dose to the target volume with a lower radiation dose to OARs as compared with photons. In cancers where radiation is part of a curative treatment strategy, protons are likely to prove significantly better than photons for a certain fraction of the patients. These subgroups must be identified from computer-based studies in which the dose distributions that can be obtained with the new technique are simulated and compared with the current standard in the same cohort of patients. These kinds of studies are referred to as *planning comparative studies* or *in silico studies*.

STEP 3: Estimation of the clinical benefit

The final step will be to determine to what extent the optimized dose distributions will translate into a clinically relevant beneficial effect. To analyse this, the dose distribution results of the individual ISPC studies are used as input for the existing NTCP models. This step is required as similar reductions in the most relevant dose distribution parameters will not always translate into similar reductions in NTCP values as illustrated in Figure 1, chapter 3. The ultimate NTCP value reduction that can be obtained depends on the shape of the NTCP curve and

the initial value of the dose distribution parameter as obtained with the reference technique.

To determine the potential benefit of IMPT in comparison with IMRT, we performed a number of ISPC studies and used the 3-step methodology to translate the obtained differences in the dose distributions into potential clinical benefits. We decided to use this study design as this methodology is one of the first and necessary steps in the development of emerging radiation techniques aiming at prevention of radiation-induced side effects. Moreover, this methodology has been adopted by the Dutch Health Council and CVZ (College voor zorgverzekeringen, the Dutch Health care insurance board) to establish the indication for proton therapy in individual cases [36].

In silico planning comparative studies

In chapter 3, we reviewed the ISPC studies performed by other investigators, comparing protons with photons in HNC. In order to value these studies, we defined a number of requirements with regard to the design of ISPC studies.

First, the new technique should be compared with the best currently available photon technique. For HNC, we considered parotid-sparing IMRT as the current standard, as a number of non-randomized and randomized studies showed that IMRT aiming at sparing the salivary glands, provides a significant reduction of the incidence of hyposalivation, physician-rated xerostomia and patient-rated xerostomia [95,134,186]. Therefore, the IMRT technique aiming at sparing the salivary glands is considered to be the current standard in head and neck radiotherapy.

Second, the gross tumour volume (GTV), clinical target volumes (CTVs) and planning target volumes (PTVs), as well as the associated prescribed radiation doses and fractionation schedules, should comply with current common practice. The studies included in the review did not always fulfil this requirement, as in some studies, only elective nodal areas were taken into account without applying a boost

to the high risk areas [40,170]. In other studies, only the high-risk areas (boost volume) were taken into account [31,126].

Third, the planning objectives with regard to target coverage and corresponding plan acceptance criteria, e.g. the recommendations of the International Commission on Radiation Units and Measurements Report 83 [83] should be clearly defined. In some studies, these objectives were inadequately described or not described at all.

Fourth, relevant targets and OARs should be delineated in a consistent way. Various studies showed significant differences between different observers regarding the delineation of both targets and OARs [37,64,78,96,151,156]. Treatment plans based on these different delineated structures may subsequently result in differences in predicted tumour control or OAR-sparing. Delineation guidelines may help to minimize inconsistencies in volume definition among radiation oncologists [195]. Although important, not all studies included in the review clearly described the way in which the target volumes were defined. Moreover, none of the studies described the way in which the OARs were delineated.

Fifth, the study should not only report results averaged over a patient group, but also the variation in the results in order to show what proportion of patients may benefit from the new technique and to what extent. Moreover, the calculation of differences in NTCP values can only be performed for individual patient data as the NTCP values not only depend on the dose reduction, but also on the absolute dose, the shape of the NTCP curve and the initial dose distribution obtained with IMRT.

Sixth, another important prerequisite for a proper design of ISPC studies is the definition of appropriate endpoints, i.e. the most relevant dose-volume parameters following from NTCP-modelling studies, and to use these parameters with properly chosen dose constraints for treatment planning optimisation for all techniques included in the analysis. Subsequently, attempts should be made to optimize the dose to the relevant OARs corresponding to these endpoints as much as possible, such that the results display the best achievable results. Only in this way it will be

possible to adequately determine which technique is best to spare the specific OARs and achieve the best results regarding the specific endpoint.

Finally, it has been recognised that, in addition to xerostomia, swallowing dysfunction is an important side effect that adversely affects health-related quality of life after radiotherapy in the head and neck area [47,105]. Therefore, sparing of the structures related to both side effects, including the parotid and submandibular salivary glands, the pharyngeal constrictor muscles and the supraglottic larynx [34,47,86,95,128] is important. Most of the published ISPC studies only attempted to reduce the dose to the parotid glands (chapter 3), whereas none of the studies attempted to reduce the dose to other relevant salivary glands [86,128] and only a few studies took into account some anatomical structures involved in radiation-induced swallowing dysfunction.

The ISPC studies in the current project met all these requirements. For the photon radiation technique we used the IMRT technique as used at the department of Radiation Oncology of the UMCG, using a simultaneous integrated boost technique using target volumes and dose prescriptions that are internationally considered current standard. The planning objectives and plan acceptance criteria were well defined. To assure consistent delineation of the target and OAR volumes we used existing delineation guidelines for the target volumes [72,73] and developed guidelines for the OARs potentially related to the most relevant radiation-induced side effects in head and neck cancer: xerostomia (chapter 2) and swallowing dysfunction [33]. Furthermore, in addition to average results, we reported the results achieved for individual patients. In the first ISPC studies, we attempted to reduce the dose to the parotid and submandibular salivary glands as much as possible without compromising target coverage. During the project, when it became clear which anatomical structures were most relevant for radiation-induced swallowing dysfunction [34], we started an additional ISPC study in order to investigate if the dose to the swallowing organs at risk (SWOARs) could be reduced by using protons without compromising the previously set treatment planning objectives (i.e. sparing of the salivary glands and a satisfactory target coverage).

Timeline: optimizing treatment planning

Figures 1 and 2 present timelines that display the progression made during this project with regard to optimization of the different treatment planning techniques for oropharyngeal cancer. Figure 1 displays the achievements made regarding the dose reductions in the OARs, while Figure 2 displays the corresponding reductions in NTCP as predicted by existing NTCP models for salivary dysfunction and swallowing dysfunction [34,86,128,164].

Salivary gland sparing

In order to illustrate the achievements made in the last 15 years in head and neck radiotherapy, we decided to start from scratch with 3D-CRT plans for ten oropharyngeal cancer cases and subsequently compared the results obtained with 3D-CRT with those obtained with IMRT and IMPT.

The first step (Figures 1 and 2) was to compare 3D-CRT (the past standard) with IMRT (the current standard). The ISPC study described in chapter 3 showed that IMRT resulted in a marked reduction of the dose to the parotid glands and to a lesser extent of the dose to the submandibular glands. Indeed, the results of a number of randomized controlled trials (RCTs) confirmed that IMRT was superior to conventional radiation techniques with regard to salivary dysfunction and xerostomia, and thus clinically validated that sparing the parotid glands indeed results in significant reductions in hyposalivation and xerostomia [18,95,134,146,186].

Pow et al. [146] compared parotid-sparing IMRT with 2D radiotherapy in nasopharyngeal cancer patients and showed that IMRT substantially improved the salivary flow rates and QOL after radiotherapy. The mean salivary flow rates were significantly higher in the IMRT arm, while in both arms QOL scores improved continually over time, but to a greater extent in the IMRT arm.

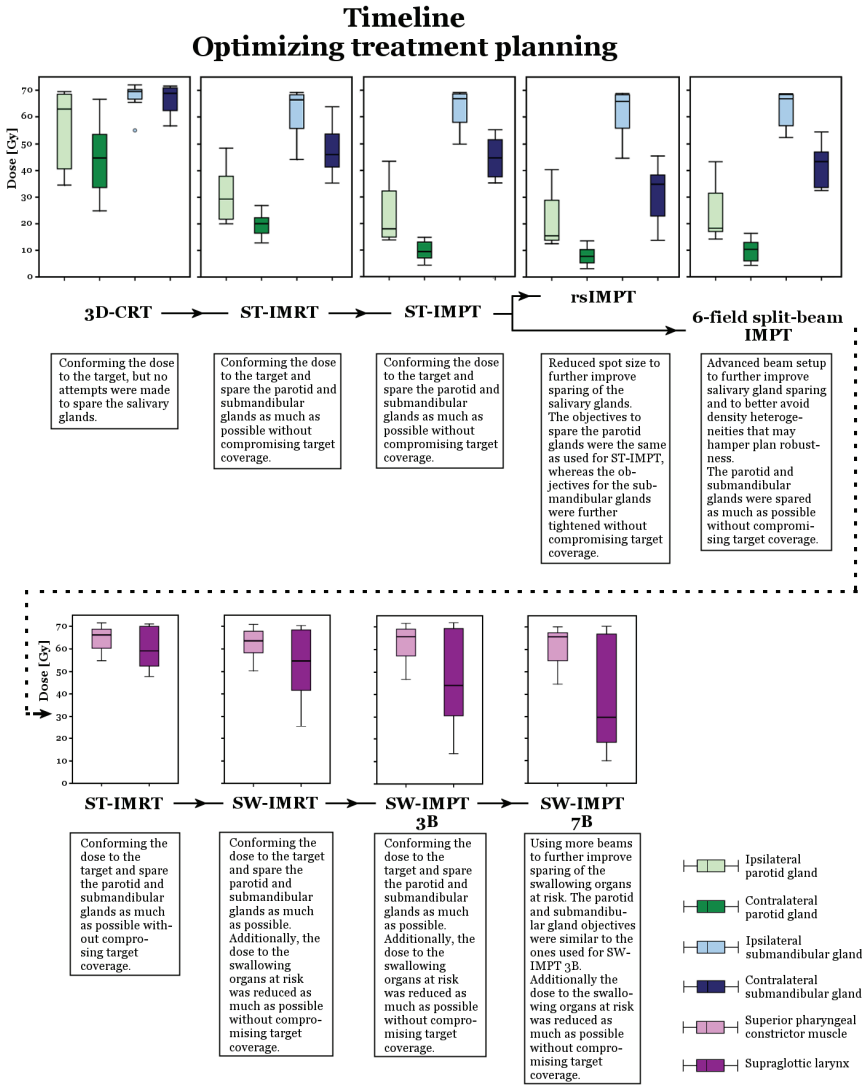

Figure 1. The progression made in this thesis with regard to optimization of the different treatment planning techniques for head and neck cancer. We started with salivary gland sparing techniques and subsequently continued with swallowing-sparing techniques. For all cases (in total 10 for the salivary gland sparing techniques and 25 for the swallowing organs at risk sparing techniques) the data is presented in box plots. *Abbreviations*: 3B, 3-beam; 3D-CRT, 3-dimensional conformal photon radiotherapy; IMRT, Intensity-Modulated Radiotherapy with photons; IMPT, Intensity-Modulated Proton Therapy; rsIMPT, reduced spot Intensity-Modulated Proton Therapy; ST, standard; SW, swallowing-sparing.

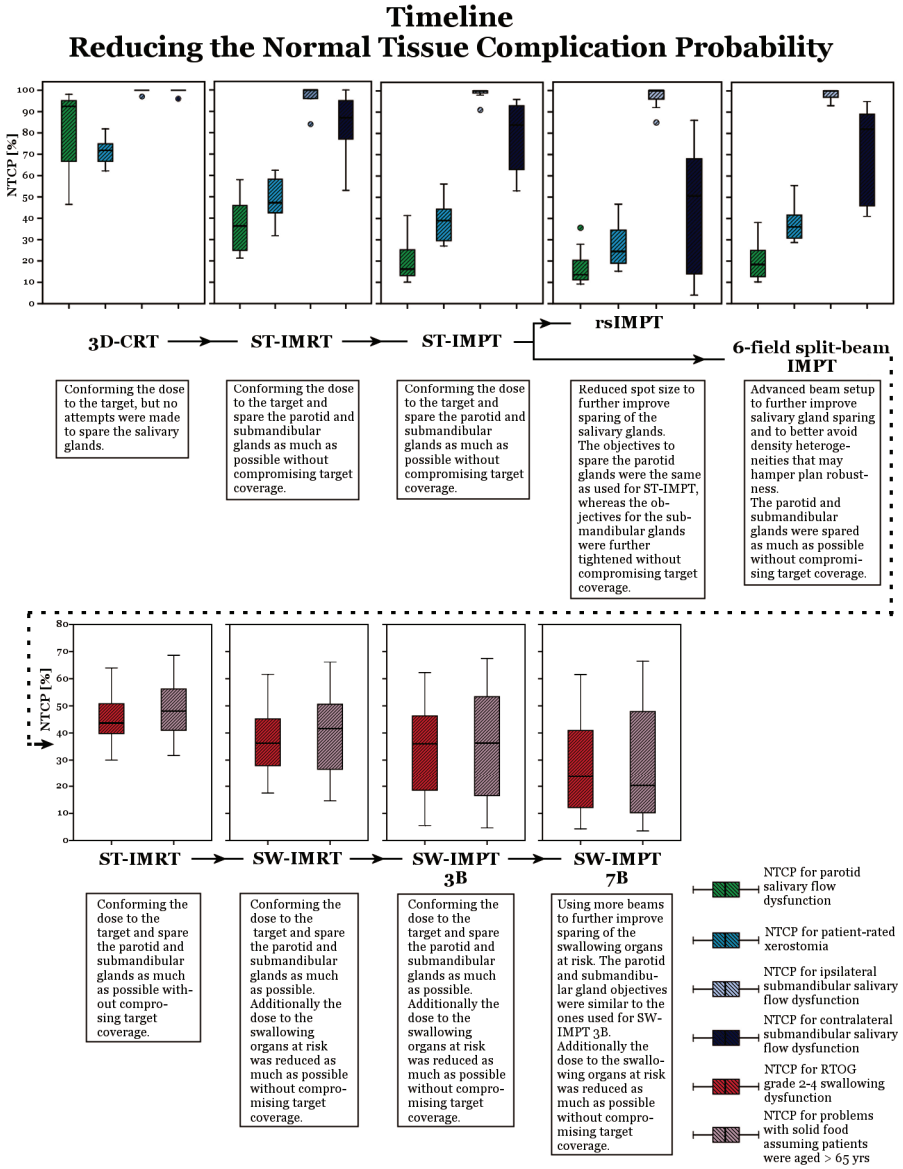

Figure 2. Normal tissue complication probabilities (NTCPs) corresponding to the dose distribution results presented in Figure 1. For all treatment planning techniques the NTCP, estimated for the studied patients by using existing NTCP models, are presented in box plots. *Abbreviations:* 3B, 3-beam; 3D-CRT; 3-dimensional conformal photon radiotherapy; IMRT, Intensity-Modulated Radiotherapy with photons; IMPT, Intensity-Modulated Proton Therapy; rsIMPT, reduced spot Intensity-Modulated Proton Therapy; RTOG, Radiation Therapy Oncology Group; ST, standard; SW, swallowing-sparing.

Kam et al. [95], also compared 2D radiotherapy with parotid-sparing IMRT in nasopharyngeal cancer patients and found similar results, i.e. significant higher salivary flow rates and lower incidences of observer-rated severe xerostomia in the IMRT arm. More recently Nutting et al. compared parotid-sparing IMRT with 3D-CRT in oropharyngeal and hypopharyngeal cancer patients [134]. This study showed that IMRT significantly reduced the incidence of xerostomia and improved saliva secretion recovery and QOL measures associated with xerostomia.

Vergeer et al. [186] performed sequential prospective cohort studies with historical comparisons among HNC patients treated with 3D-CRT and IMRT subjected to a prospective standard follow up program. They showed that parotid-sparing IMRT compared with 3D-CRT resulted in a significant reduction of patient-rated and observer-rated xerostomia [186]. Moreover, this study also indicated that IMRT reduced other head and neck symptoms (like trismus and head and neck pain), which eventually translated into significant improvements of the more general dimensions of health-related QOL.

In summary, the results of these RCTs showed that the dose reduction to the salivary glands obtained with IMRT compared with 2D- or 3D-CRT indeed significantly reduced the risk of salivary dysfunction and/or xerostomia and thus confirmed that the results of previous ISPC studies comparing IMRT with 3D-CRT translated into the expected clinical benefit.

The second step in the current project (Figures 1 and 2) was to compare 3D-CRT (the past standard) and IMRT (the current standard) with scanned 3-field IMPT. Protons can be delivered by two techniques: passive scattering and the more advanced active scanning technique (as previously explained in chapter 1). In the past, most institutes only used treatments with scattered protons. This technique is limited by the homogeneous dose delivered to the target for each single field and by the fixed length of the Spread-Out Bragg Peak (SOBP), resulting in an increased dose proximal to the target (see chapter 1, Figure 3 of this thesis). In contrast, scanned proton therapy allows for a non-uniform dose distribution to be delivered for each field (IMPT) and hence for more degrees of freedom with regard to target

coverage and OAR-sparing (chapter 1, Figure 4). Until recently the active scanning technique, was only available at the Paul Scherrer Institute (PSI) in Switzerland (since 1996).

Nowadays, more proton therapy facilities with scanned proton therapy have been developed or are under construction [139]. Consequently, scanned IMPT will become more widely available. The availability of active scanning techniques is of importance as the results of the various ISPC studies discussed in our review study (chapter 3) suggest that scanned IMPT provides more optimal results with regard to OAR-sparing while maintaining adequate target coverage than scanned 3-dimensional conformal proton therapy (3D-CPT) and scattered 3D-CPT.

Results of the comparison between IMRT and 3-field IMPT – from now on referred to as standard (ST) IMPT – showed that ST-IMPT allowed for a significant reduction of the mean dose to the parotid glands, but not to the mean dose to the submandibular glands. More specifically, the contralateral and ipsilateral parotid gland dose reduction was 5 Gy or more in 90% and 80% of the cases, respectively. The dose reductions that could be obtained to the contralateral submandibular gland were much lower (5 Gy or more in 30% of the cases) and it was generally impossible to reduce the dose to the ipsilateral submandibular glands.

Assuming an NTCP reduction of 10% as clinically relevant, the estimated percentage of patients that would benefit form ST-IMPT were 80% and 40% for severe salivary flow decline and patient-rated xerostomia, respectively. It is important to note that both the parotid and the submandibular gland doses are associated with patient-rated xerostomia [86]. Therefore, in addition to parotid gland sparing, submandibular gland sparing is also of major importance.

As discussed above, the first ISPC study presented in this thesis (chapter 4) showed that ST-IMPT, as compared with IMRT, only allowed for relatively small dose reductions to the submandibular glands. Moreover, in 80% of the patients, the mean dose to the ipsilateral submandibular gland was somewhat higher with ST-IMPT than with IMRT (Figure 4 chapter 4).

There are two possible explanations for these findings. First, the lateral dose fall-off (the penumbra that affects the possible dose gradient steepness) for protons was larger than for photons. Compared to the parotid glands, the submandibular glands are substantially smaller and overlapped relatively more with the planning target volume (PTV). Thus, the absolute submandibular gland volume located outside the PTV that could be spared was generally smaller than the absolute parotid gland volume outside the PTV. Hence, steeper dose gradients than currently used, are necessary to improve sparing of the non-target (non-overlapping) submandibular gland part. It seems that for these particular cases, the advantage of the steep distal fall-off at the end off the Bragg peak is not sufficient to improve sparing of the submandibular glands. Apparently, the width of the lateral penumbra of the proton beam is also important for the optimization of the dose reductions in these small OARs that overlap with the planning target volume. Second, inhomogeneities such as bone-soft tissue and/or soft tissue-air transitions along the proton beam path may have influenced the results. For the standard IMPT plans, 3 fields were used as a previous study showed that target dose homogeneity and OAR-sparing were similar with 3-, 5- and 9-field IMPT [170]. It should be emphasized, that with ST-IMPT the anterior-lateral proton beams (50 – 60°) propagate through the oral cavity region. These beams are thus possibly affected by density inhomogeneities, i.e. bone-soft tissue and/or soft tissue-air transitions (density inhomogeneity effects are taken into account by the used proton pencil beam dose calculation algorithm [143,158]). The effects are degradation of the sharp distal edge of the Bragg peak and variable ranges of the individual protons, resulting in degraded dose distributions [66,115]. Thus, proton beams that, due to their orientation in relation to the patient's anatomy, traverse inhomogeneities before stopping near the submandibular glands contribute to degradation of the dose distribution in that region.

As it is considered worthwhile to spare the submandibular glands in addition to the parotid glands and to improve avoidance of areas that are prone to variations in density heterogeneities, the next step was to investigate if more advanced IMPT

techniques could further optimize the dose distributions (Figure 1). For that purpose, we explored two different advanced IMPT strategies and compared the results with the results as obtained with standard 3-field IMPT.

The first adaptation was the introduction of a reduced proton beam spot size. For simulation of the standard 3-field IMPT plans a relatively wide proton pencil beam (lateral width) was assumed. However, the new generation of scanning proton systems allow for proton beam scanning with a smaller spot size. As smaller proton beam sizes may result in steeper dose gradients, these advanced scanning proton systems may result in more conformal target doses and allow for improved sparing of relevant OARs. To determine whether this hypothesis was correct, we investigated if IMPT with a reduced spot size (rsIMPT), compared with ST-IMPT, could further improve parotid and submandibular gland sparing (Figure 1). This study (chapter 5) showed that 3-field rsIMPT significantly reduced the dose to both the parotid and submandibular glands, yielding a potential clinical benefit with regard to both parotid and submandibular gland salivary flow dysfunction and patient-rated xerostomia (Figure 2).

It should be remarked that the dose in the parotid glands was reduced automatically, without further tightening of the parotid gland dose-volume objectives (DVOs). Moreover, we made no efforts to further tighten the parotid gland DVOs with rsIMPT compared with the standard IMPT plans, as the parotid gland dose was already relatively low with ST-IMPT. The mean contralateral and ipsilateral (overlapping relatively more with the PTVs) parotid gland doses varied from 4.3 Gy to 14.9 Gy and from 13.9 to 43.4 Gy, respectively, whereas the mean contralateral and ipsilateral mean submandibular gland doses were generally much higher varying from 35.2 to 55.2 Gy and 49.9 to 69.2 Gy, respectively. Therefore, further efforts were only made to reduce the submandibular gland doses as much as possible (by tightening of only those DVOs) and thus further improvement with regard to parotid-sparing can not be excluded. The results showed that, in contrast to ST-IMPT, the doses to the contralateral submandibular gland could be further reduced with rsIMPT. Similar results were obtained for the ipsilateral submandibular gland dose. Where ST-IMPT compared with IMRT yielded a dose

reduction of at least 5 Gy in only 30% of the cases for the contralateral submandibular gland and in none of the cases for the ipsilateral gland, these percentages were 100% and 20%, respectively, when rsIMPT was used. Furthermore, rsIMPT substantially increased the number of cases with a contralateral submandibular gland mean dose below 39 Gy, which has been found to be the dose threshold [128] that allows for a submandibular gland flow rate recovery. With IMRT, the mean dose to the contralateral submandibular gland could be kept below 39 Gy in only 20% of the cases, while with ST-IMPT and rsIMPT, these rates were 30% and 80%, respectively.

Clinical studies that investigated the feasibility of submandibular gland sparing with radiotherapy showed that sparing of those glands resulted in a reduced probability of xerostomia [111,155]. Saarilahti et al. [155] reported that submandibular gland sparing IMRT (sparing only the contralateral gland), compared with non-submandibular gland sparing IMRT, resulted in a substantially improved unstimulated saliva flow, i.e. 60% referenced to baseline flow versus 25% referenced to baseline. This outcome translated into a significant reduction of the incidence of grade 2-3 xerostomia from 61% without sparing the submandibular gland to 22% when the submandibular glands were spared ($p=0.018$) [155]. Liu et al. compared a test group consisting of patients undergoing surgical transfer of one submandibular gland (to the submental space) before radiotherapy, which is subsequently shielded during radiotherapy, with a control group in which no submandibular gland transfer took place [111]. Surgical transfer of the submandibular gland substantially improved saliva excretion, i.e. the mean saliva weight was 1.65 g in the surgical transfer group versus 0.73 g in the control group ($p<0.0001$). In addition, surgical transfer reduced the incidence of moderate-to-severe xerostomia from 78.6% in the control group to 12.9% in the surgically transferred group [111]. In these latter studies, submandibular gland sparing was achieved by compromising the target coverage [155] or by surgical transfer of the submandibular glands up-front of radiotherapy [111]. The present ISPC study, however, demonstrates that rsIMPT as compared with IMRT and ST-IMPT, may result in an additional reduction in the contralateral submandibular gland dose

below the threshold dose, without compromising target coverage or surgical transfer of this submandibular gland. It should be noted that submandibular gland sparing with the surgical transfer method will not be applicable for all patients [48]. Moreover, in many patients with HNC, level IB (including the submandibular gland) is part of the (prophylactic) target volume, and cannot be spared without a major compromise to the target volume. So, submandibular sparing is only relevant in those cases in which level IB is not part of the target volume, e.g. in some cN0-cases. Furthermore, when the primary tumour is too close to the submental space, the transferred gland cannot even be shielded during radiotherapy.

The second adaptation we made to improve submandibular gland sparing concerned the geometry of the proton beams. In the ST-IMPT a standard setup was used with beams that traverse the oral cavity and the lung apex. However, the traversed density heterogeneities highly degraded the proton Bragg peak resulting in variable ranges of the individual protons [115] and thereby in less steep dose gradients that are of importance regarding target dose conformity and OAR-sparing. We hypothesize that improved avoidance of those density heterogeneities that degrade the dose distribution, will help to improve sparing of the small submandibular gland volumes that always overlap with the PTVs. In the head and neck area density heterogeneities are common and the size and shape of some may even vary from day to day. The anatomic structures in the skull base area are less prone to density heterogeneity variations. However, the more caudal parts of the head and neck area, comprising the oral cavity, including the mobile tongue and air cavities, is more sensitive to density heterogeneities that may vary from day to day. Thus, in addition to the effect of static density heterogeneities (no daily variations), density heterogeneities that vary from day to day or even during treatment may affect the actually administered dose distribution on that specific day. In this regard, avoiding the oral cavity may improve treatment plan robustness, which refers to the sensitivity of a plan to for instance anatomical changes during treatment resulting from weight loss, tumour shrinkage, radiation-induced changes

to normal tissues and/or movements of structures such as the tongue. This issue will be further discussed in the paragraph “Future directions”.

Therefore, the next step (Figure 1) was to investigate whether a more advanced 6-beam split-field IMPT technique could further improve two aspects of the dose delivery, that is: (1) reduction of the radiation dose to the parotid and submandibular glands, and (2) avoidance of those air cavities and other structures that may cause tissue heterogeneities resulting in a less accurate dose delivery. The results showed that 6-beam split-field IMPT, compared with ST-IMPT, significantly reduced unnecessary normal tissue irradiation (non-target tissue, e.g. including the oral cavity, the sublingual glands and the mandible body) and improved sparing of the contralateral submandibular gland, whereas similar dose results were obtained for the ipsilateral submandibular gland and the parotid glands (chapter 6).

Compared with ST-IMPT, 6-beam split field IMPT resulted in 80 % of the cases in a reduced mean contralateral submandibular gland dose (ranging from 0.2 to 4.6 Gy) and therefore in a reduction of NTCP values for salivary flow dysfunction (Figure 2). In 40% of the patients the estimated NTCP reductions were at least 5% and in two cases, NTCP reductions up to 16 and 17% were obtained. Thus, when an NTCP reduction of 10% was defined as clinically relevant, 20% of this study cohort would benefit from 6-beam split-field IMPT regarding contralateral submandibular salivary flow dysfunction.

Where ST-IMPT compared with IMRT yielded a dose reduction of at least 5 Gy in 30% of the cases for the contralateral submandibular gland and in none of the cases for the ipsilateral gland, these percentages were 40% and 20%, respectively, when 6-beam split-field IMPT was used

Although no specific constraints were applied to the sublingual glands, the oral cavity and the mandible, 6-beam split-field IMPT significantly improved sparing of these structures in almost all cases (depending on the OAR), indicating an improved avoidance of the oral cavity region that may encompass variable density heterogeneities, as compared with ST-IMPT. Hence, this beam geometry

may not only result in lower dose to the submandibular gland but also contribute to improved robustness of proton therapy in HNC.

With 6-beam split-field IMPT the ipsilateral submandibular gland dose was generally similar to those obtained with ST-IMPT. This can be explained by the larger width of the lateral penumbra of the proton beam with 6-beam split-field IMPT compared with that with rsIMPT. Another explanation could be that in general the ipsilateral submandibular gland part outside the PTV is substantially smaller compared with that of the contralateral gland. Only 20% (range: 0-54%) of the ipsilateral gland volume was located outside PTV1, compared to 69% (range: 38-82%) of the contralateral gland volume. In addition, the ipsilateral gland often substantially overlapped with PTV2 (the boost target). Therefore, much steeper dose gradients are necessary to spare the ipsilateral gland outside the PTV compared to those necessary for the contralateral gland. Furthermore, the extra lateral proton beam (±70°) in the 6-beam split-field IMPT plans (only present on the ipsilateral site to assure adequate PTV2 coverage) also adds an extra dose to the ipsilateral submandibular gland.

As the new beam geometry predominantly used posterior beams that may have to traverse the OARs involved in swallowing dysfunction (SWOARs), the dose to the pharyngeal constrictor muscles and larynx structure were analysed as well. However, no increase of the dose to these structures was observed with the 6-beam split-field setup.

Swallowing organs at risk sparing

The number of patients that are currently treated with more intensified treatment regimens such as concomitant chemoradiation is increasing, resulting in more acute and late side effects [62]. Consequently, the incidence of swallowing dysfunction is gradually increasing. Swallowing dysfunction has a major impact on health-related quality of life as reported by patients, and appears to be even more important than radiation-induced xerostomia [105]. Therefore, dose reductions to the anatomical structures involved in swallowing may further improve patient's quality of life.

An important prerequisite for optimising radiotherapy treatment planning to reduce swallowing dysfunction is knowledge with regard to which dose-volume parameters in which anatomical structures are most relevant for the development of swallowing dysfunction. This was the main purpose of another project recently performed at our department [34]. Based on a literature review, it was concluded that the available data did not allow for a clear definition of the most relevant DVH-parameters and for identification of the most important organs at risk [35]. Similar to what was done for the current project, guidelines for unambiguous contouring of the candidate SWOARs were defined and published [33]. Subsequently, a prospective cohort study was performed in order to develop a predictive model for swallowing dysfunction after curative (chemo)radiation. The primary endpoint was physician-rated RTOG grade 2-4 swallowing dysfunction as assessed 6 months after completion of radiation therapy. The secondary endpoints were moderate to severe patient-rated problems with respect to swallowing solid, soft or liquid food and choking when swallowing, all as assessed at 6 months after the end of treatment. The most relevant and predictive SWOAR dose-volume parameters were the pharyngeal constrictor muscle (PCM) superior mean dose, the supraglottic larynx mean dose, the middle PCM mean dose and the oesophagus inlet muscle (EIM) receiving at least 60 Gy, EIM $V_{60\ Gy}$ (in this order of priority) [34].

A number of studies already investigated the potential benefit of IMRT with regard to sparing the SWOARs as compared with standard IMRT, only aiming at reduction of the mean dose to the parotid glands. Eisbruch et al. [53] indicated that sparing of the pharyngeal constrictor muscles and larynx with IMRT, compared with standard IMRT, resulted in a 10% and 7% reduction of the $V_{50\ Gy}$ (the volume receiving at least 50 Gy) for the pharyngeal constrictor muscles and the larynx, respectively. Some other authors investigated the potential of IMRT to spare swallowing organs by using a split-field IMRT technique [27,63,190]: IMRT fields that irradiated the superior target were matched to an anterior low-neck field that blocked part of the pharynx and oesophageal region. While this split-field technique reduced the dose to the inferior part of the SWOARs (like the inferior pharyngeal constrictor muscle, the larynx and part of the oesophagus), the dose to

the inferior part of the target volume was reduced below the minimally required dose. Additional to these studies, in a more recent study we investigated the potential benefit of swallowing-sparing IMRT (SW-IMRT) compared with standard IMRT (only sparing the parotid and submandibular glands) in various head and neck tumour sites [183]. In that study, efforts were made to spare the most relevant SWOARs, as previously described [34]. Additional sparing of the SWOARs with SW-IMRT was possible without increasing the salivary gland doses or violating other plan acceptance criteria such as adequate target coverage. However, it appeared that SW-IMRT resulted in an increased volume (the average increase was 41 cm^3 [183]) receiving 95% of the prescribed PTV1 (lymph node regions) dose, indicating a worsening of the target dose conformity. These results support efforts to improve the SWOAR-sparing technique in HNC patients with new techniques such as scanned proton therapy.

Therefore, the next step was to perform an ISPC study comparing SW-IMRT with swallowing-sparing IMPT (SW-IMPT) in a patient data set existing of 21 oropharyngeal and 4 hypopharyngeal cancer cases (chapter 7). In addition to standard (ST) IMRT and ST-IMPT plans, for each patient one swallowing-sparing (SW) IMRT plan and two SW-IMPT plans with a 3-beam and a 7-beam setup, respectively, were created (Figure 1). In the standard plans the parotid and submandibular glands were spared as much as possible, whereas in the SW-IMRT and SW-IMPT plans, additional objectives were applied to the relevant SWOARs (as specified above [34]).

Compared with ST-IMRT, the mean contralateral and ipsilateral parotid gland doses were lowest with the ST-IMPT and SW-IMPT plans in all cases. Compared with ST-IMRT, in 52% of the cases IMPT reduced the contralateral submandibular gland dose and in 7 cases those reductions were at least 5 Gy (dose differences ranged from -4.4 up to 29.2 Gy, including all 25 cases).

With SW-IMRT, the dose to the SWOARs was reduced in relation to ST-IMRT, while the lowest SWOAR doses were obtained with SW-IMPT with a 7-beam setup. Compared with the current standard, ST-IMRT, supraglottic mean dose reductions of at least 5 Gy could be obtained in 44% of the cases with SW-IMRT, in 60% of the

cases with 3-beam SW-IMPT and in 68% of the cases with 7-beam SW-IMPT (chapter 7).

According to the recently developed NTCP models for physician-rated and patient-rated swallowing dysfunction [34] these dose reductions translate into markedly reduced probabilities for different aspects of swallowing dysfunction. For both the SW-IMRT and SW-IMPT techniques, the largest estimated NTCP value reductions were found for RTOG grade 2-4 swallowing dysfunction and patient-rated moderate to severe problems with swallowing solid food. The largest NTCP value reductions were again obtained with SW-IMPT with a 7-beam setup (Figure 2). Reference to ST-IMRT, and assuming a 10% reduction in NTCP as clinically relevant, it was predicted that 64% of the patients would benefit from 7-beam SW-IMPT to prevent RTOG grade 2-4 swallowing dysfunction, whereas these percentages were 52% and 40% with 3-beam SW-IMPT and SW-IMRT, respectively. As compared to SW-IMRT (instead of ST-IMRT), 48% and 28% of the cases would benefit from 7-beam SW-IMPT and 3-beam SW-IMPT, respectively, demonstrating the potential clinical relevance of SW-IMPT.

When further analysing the results of this study, some important issues came to light. First, in contrast to the different IMPT plans, SW-IMRT target conformity became worse when attempts were made to spare the SWOARs in addition to the salivary glands. The volume receiving 95% of the PTV1 prescription dose increased, similar to the results shown in a previous study [183]. Second, on average, the mean supraglottic dose, the parotid and contralateral submandibular gland doses and the volume receiving a low-to-intermediate dose were significantly smaller with IMPT. These results indicate that overall, IMPT allows for (1) improved target conformity, (2) better sparing of the salivary glands (3) better sparing of the SWOARs and (4) smaller volumes receiving a low-to-intermediate dose.

It is important to notice that the benefit of swallowing-sparing radiation techniques varied widely among individual patients. This case dependency is due to variation in normal patient anatomy and to variation in size and extent of the overlap between OARs and PTVs. The patient population used in the current project was predominantly composed of oropharyngeal cancer patients and the

primary tumour frequently overlapped with the PCM (see for instance Figure 1 chapter 7), which made it sometimes hard or even impossible to spare the SWOARs. In contrast, in most cases, the supraglottic larynx was located more distant from the target and could therefore be spared more easily. These results illustrate that the selection of patients for proton therapy instead of photon therapy can be done most optimally by performing ISPC studies for each individual patient. In this regard the 3-step methodology described presents one of the most powerful examples of individualized treatment. This is of major importance as the currently available literature does not allow for a clear distinction (e.g. based on tumour characteristics in relation to the location of the relevant OARs) between patients who will substantially benefit from protons and those who will not, without having information on the actual and ultimate dose distributions. Further research will focus on this subject.

In summary, the results of the current project indicated the potential benefits of IMPT compared with the currently used most advanced photon IMRT techniques in head and neck cancer therapy. It was shown that IMPT is expected to result in significantly reduced risks of the two most important radiation-induced side effects, xerostomia and swallowing dysfunction. Individual planning comparative studies are required to select patients that will most likely benefit from protons instead of photons.

Future directions

Clinical studies in which protons were used to treat HNC are scarce and if performed, only partly used protons during the radiation course [26,28,29,167,173,176]. Therefore, these clinical studies did not fully explore the benefits of protons regarding the potential to improve normal tissue sparing or target dose escalation without increasing normal tissue doses. In the current project, we applied protons for the full treatment course that showed that protons have the potential to reduce the dose to clinically relevant organs at risk, without

reducing the dose to the target volume. In this respect, the 3-step approach as described in this thesis can be regarded as hypothesis generating, which however requires further validation in clinical trials.

Clinical validation

The clinical validation of protons aiming at reduction of side effects as compared with photons can be done in two different ways, including: (1) participation in randomized controlled trials (RCTs), or (2) treatment with proton therapy within the framework of the so-called sequential prospective observational studies (SPOS) with historical comparisons.

Currently, evidence-based medicine has become the cornerstone in the development of radiation oncology and RCTs are still considered the gold standard for determining improvements in clinical outcome between different strategies, including new technologies such as protons. However, currently, the results of RCTs comparing protons with photons are lacking for several (good) reasons as pointed out by Glimelius et al. and Goitein et al. [65,70]. There is no doubt that whenever possible, an RCT is the preferred methodology to clinically validate the added value of new treatment modalities, including that of proton therapy, regarding tumour control and survival as well as acute and late radiation-induced side effects. However, it has been stated that new technologies aimed at improving treatment quality should primarily be tested using process measures or operational characteristics, as the sensitivity and specificity of clinical outcome is low for detecting quality improvements [12]. In the case of proton therapy, there may be circumstances that an RCT is not considered feasible, e.g. in case a side effect with a major impact on QOL (e.g. visual impairment) is likely to be prevented by protons compared with photons without compromising the dose to the target volumes or in case of a major estimated NTCP reduction of a side effect that interferes with QOL (e.g. swallowing dysfunction), even without compromising the dose to the target volume. For these cases, an alternative methodology, the so-called SPOS has been invented [77]. In a SPOS, patients treated with photons and protons are subjected to similar standard follow up programs in which relevant baseline and treatment

characteristics, acute and late radiation-induced morbidity and patient-rated QOL are determined in a well structured standardised prospective program. Preferably, similar prospective observational studies should already be performed in patients treated with photons before proton therapy is available to avoid selection bias. In such an approach, the results obtained in a population treated with the current standard (e.g. photon IMRT) can be considered as the reference for those who will be actually treated with protons. Moreover, this reference population can be used for the statistical power analysis in order to estimate the required number of patients that has to be treated with protons to show a significant benefit of the new technique. In such an approach, in each individual patient, the dose distributions obtained with the best available photon technique has to be weighted against those obtained with the best possible proton technique.

Irrespective of the design chosen for the clinical validation of protons aiming at reduction of side effects, selection of the proper patients is of major importance. As pointed out previously in this discussion, the expected NTCP value reduction of protons compared to photons can be estimated using the 3-step methodology. This 3-step approach is necessary to ensure proper patient selection, as it does not make any sense to include those patients in which the individual 3-step approach already indicate that they will not benefit from protons at all. In particular when this category of patients is relatively high, there is a major risk of the so-called dilution effect increasing the probability of false negative results of such clinical validation study, either an RCT or an SPOS. In this respect, the eligibility criteria of any clinical validation study should at least include the results of an ISPC study and subsequent estimation of the NTCP value reduction of the side effect that has been appointed as primary endpoint, in each individual patient. It should be stressed that such an approach should always be part of any selection procedure for proton therapy.

In such an approach, the NTCP model obtained in patients treated with photons should be validated in those treated with protons (this subject will be discussed in more detail in the paragraph “Interpretation of the NTCP outcome results”). If the NTCP models are similar and the observed reduction of the

incidence of the side effect investigated is equal or higher than the expected reduction, this will more strongly support the benefit of protons over photons.

Currently, the number of proton centres that actually treat head and neck cancer patients with a full course of spot scanning proton therapy is very small. This is mainly due to the fact that the number of proton centres that actually have the possibility to use spot scanning is very limited. Moreover, the clinical application of spot scanning proton therapy requires a number of measures in order to ensure that its application can be carried out safely. These measures will be discussed in more detail.

Image-guided radiotherapy (IGRT)

As already pointed out before, delineation guidelines for consistent definitions of all relevant structures, including target volumes and OARs, are extremely important in order to improve reporting and interpretation of radiation treatment results across studies (chapter 1). Nevertheless, consistent definition of all relevant structures on the regular planning CT scan may not be sufficient, as during the course of fractionated radiation treatment, anatomical changes may occur that deteriorate the dose distribution. In HNC, both target volumes and normal tissue volumes may significantly change during the radiation course [8,25,185]. Such changes may result in differences between the planned and the actually delivered radiation dose [25]. These differences are expected to be larger for protons than for photons, as protons have a finite range that depends on the traversed tissue densities. In order to be able to notice these anatomic changes and adapt the treatment plan when necessary, image guidance during the entire treatment course will be of major importance. Moreover, application of image-guided radiotherapy is one of the necessary steps to ensure adequate target doses and limited doses to the healthy tissues for both photon and proton therapy.

Since protons have a limited range with a sharp distal fall-off, they are more sensitive to density inhomogeneities within the patient than photons. Interfaces in density along the beam path (e.g. tissue–bone) affect protons in a different way than photons. When a uniform slab of material of a higher density (e.g. bone) than

the surrounding tissue is present in the beam path, the photon beam's intensity (and hence the dose) just behind that slab will be reduced by a certain amount (depending on the slab's composition and thickness). In contrast, the proton beam's intensity just behind the slab would remain similar, while the proton range would be strongly affected [66]. Therefore, the entire SOBP curve is shifted in depth, which causes an underdosage at the distal part of the target, as schematically illustrated in Figure 3. In contrast, a slab of much lower density (like that of air), will result in an increased dose to structures behind the target. In case tissue densities vary in the direction orthogonal to the beam direction, the range of an individual proton depends on its' trajectory within the beam. Protons traversing a very dense region will have a lower residual range compared with protons traversing a less dense region. Hence, density heterogeneities across the proton beam spread out the Bragg peak, resulting in a degraded Bragg peak [66,115]. Due to these effects proton beams are more sensitive to variations in density heterogeneities along their path than photon beams and therefore to patient setup variations.

Daily image guidance reveals whether the patient's anatomy deviates from the anatomy as visualized at the CT scan used for treatment planning. Albertini et al. investigated the influence of substantial weight gain or loss on the dose distribution achieved with IMPT in two paraspinal cancer cases [2]. The dose differences resulting from the substantial weight changes were not as large as expected based on the resulting differences in range. A maximum proton range difference of up to +8 mm or -13 mm, resulted in an increased dose to the cauda equina (the relevant critical structure) of only 2%. Proton range differences of +15 mm and -16 mm as found for the PTV, only slightly affected target dose homogeneity. Still, the authors advised to perform repeated CT scans and to re-plan the patient in case of substantial weight changes to avoid the presence of new hot and/or cold spots within or outside the PTV.

In head and neck treatment plans the relevant anatomical structures, the beam setup, the spot distribution and spot weights are very different from the paraspinal

treatment plans. Therefore, the effects of anatomical changes in the head and neck region on the dose distribution are expected to be different from the effects as demonstrated in the paraspinal cancer cases. Additional studies are needed to investigate this in more detail. At this moment, it seems appropriate to re-scan head and neck cancer patients in case of a substantial change in the anatomy that may be of influence on the proton dose distribution.

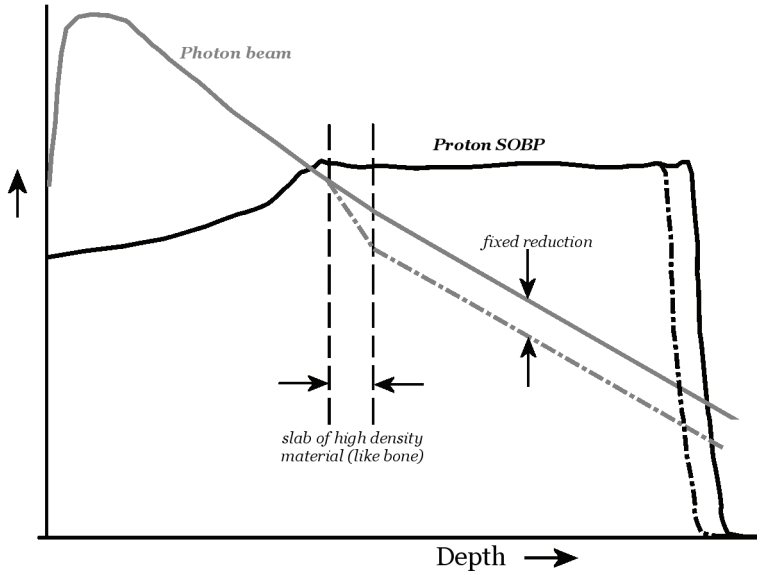

Figure 3. Schematic display of the effect of density inhomogeneities along the beam path on the depth dose curve of photons (grey lines) and protons (black lines). The solid lines display the depth dose curve when the beams irradiate a homogeneous tissue and the dashed lines display the effect of a slab of high density material within the homogeneous tissue, affecting the depth dose curves.

A possible tool for online visualization of the patient's anatomy is an X-ray tube that displays the beams eye view during treatment, as currently constructed in the new gantry developed at PSI. Those images, however, can only verify the patient's position based on bony reference structures and are not suitable to verify whether soft tissue changes occurred during the proton therapy course, e.g. to verify whether the patient gained weight or if tumour shrinkage occurred. Magnetic

resonance imaging (MRI), on the other hand, provides very good soft-tissue contrast.

In radiotherapy treatment planning, diagnostic MRI scans are often fused with the treatment-planning CT scan to improve target volume and OAR delineation. Currently, the integration of a MRI scanner with a 6 MV linear accelerator is investigated and a clinical prototype hybrid MRI linac for on-line MRI-guided radiotherapy is under construction [104,147]. This online soft tissue imaging technique will allow for visualization of online anatomic motion and for gated radiotherapy. The feasibility of MRI-guided proton therapy has been investigated as well [148]. The first results are promising [148,161] and it seems that, in contrast to photon therapy, the effect of the magnetic field on the proton dose distribution is very small [148]. Moreover, the shift of protons in a 0.5 T field is relatively small (200 MeV protons are deflected by 8 mm at the end of their 25 cm range) and this shift can be implemented in a treatment planning program [161]. Furthermore, the energy of the secondary electrons is low (in the order of keV versus ~1 MeV in case of 6 MV photons [148]), therefore their range is limited and they will have an almost negligible effect on the dose distribution. In other words, the effect of the electron return effect, caused by the Lorentz force, on the dose distribution is almost negligible [148,161]. Nevertheless, there are still several important issues that should be further investigated, including the effect of the magnetic fields of the MR on (a) the proton beam transport magnets and on (b) the proton beam monitoring system and whether the electron return effect may slightly increase the LET and RBE [148,161].

Another imaging tool that may help to improve the clinical application of proton therapy is Positron Emission Tomography (PET). Due to the nuclear reactions between the incident proton beam and the nuclei in the irradiated tissue, positron emitting nuclei are created (predominantly ^{15}O ($T_{1/2}$ = 222 s) and ^{11}C ($T_{1/2}$ = 20.3 min), having a proportion of 11/1 in activity at 3 min after treatment, respectively [189]). The emitted positrons subsequently annihilate with electrons from the surrounding tissue under emission of two photons oriented in opposite directions which can be detected by a PET scanner. These positron emitting nuclei

can be produced along the proton path until the proton energy drops below the threshold of nuclear reactions (up until the last few millimetres of the end of the proton range [66]). (This energy threshold equals 16.6 MeV and 20.3 MeV for ^{15}O and ^{12}C, respectively [189].) Consequently, PET imaging allows for visualization of the treatment dose distribution and can be used as a treatment plan delivery verification tool. Plan verification can be performed by comparing the measured positron emission activity distribution with a precalculated positron emission activity predicted by for instance Monte Carlo calculations [138]. Several investigators have explored the use of PET imaging for both off-line (using PET/CT imaging postirradiation) and on-line (using an in-room PET system) for treatment plan verification, and the results are promising [100,124,132,138]. Clinical studies with off-line PET/CT demonstrated the feasibility of post-irradiation PET/CT for treatment verification, but also indicated that there are some remaining hurdles that need to be tackled to further improve this technique, including the negative effect of (1) the biological washout of the induced positron activity (due to the time, 10-30 min, required for moving the patient to the PET room) and (2) the re-positioning of the patient [100]. These problems can be tackled by on-line PET imaging, as shown in Japan [132]. PET scans are performed each treatment day and the scan of the first treatment day is used as reference, as accurate models to predict the activity distribution are not yet available [132]. Scans of the subsequent treatment days are compared with the reference scan and when necessary (i.e. when substantial differences between the activity distributions were visualized due to for instance tumour size reduction or other changes in the patient's anatomy) a new CT is made and replanning takes place.

Currently, research regarding the clinical use of PET imaging as a plan delivery verification tool focuses on (1) in-room PET imaging and (2) further development of accurate and fast methodologies to predict the positron activity distributions [7,100,124,125].

Prompt gamma (PG) imaging, currently under development, is also a promising tool for range verification in proton therapy. Inelastic interactions of protons may excite nuclei to a higher energy state that will very quickly (10^{-19} —

10^{-9} s) return to their ground state under emission of a single photon, called a prompt gamma (with energies mainly between 0 to 7 MeV) [169]. As prompt gammas are emitted instantaneously, isotropically, and have a reasonably high production rate for a typical clinical dose of 2 Gy/min, this imaging tool seems very useful for online proton range verification [127]. This technique can be used in the same way as the previously discussed PET imaging tool: by comparing measured with expected simulated PG distributions. Monte Carlo simulations in different clinical cases, comparing PET and PG imaging for proton range verification, showed that the PG imaging method is very promising as (1) it has a ten times higher gamma production rate than the PET imaging method and (2) regarding the measured activity as a function of depth, the prompt gamma distribution was on average ~5 to 10 mm closer to the dose fall-off than the PET distribution [127]. However, PG measurement systems suitable for online range verification of proton therapy in clinical practice are not yet available. Currently, various researchers focus on the development of such PG measurement systems [121,169].

Plan robustness

Treatment uncertainties, like intra-fraction changes (e.g. due to organ movement or patient movement) and inter-fraction changes (e.g. due to anatomical changes occurring during the radiotherapy course or variations in patient positioning) can affect the actually delivered dose distribution. A robust treatment plan refers to a plan for which the planned dose distribution agrees well with the actually delivered dose distribution even in the presence of such treatment uncertainties. Usually, treatment plans are evaluated by investigating target coverage and OAR-sparing and treatment plan acceptance criteria are defined by specific DVH parameter values. Treatment plan robustness can be analysed by checking whether the plan affected by uncertainties still satisfies the treatment plan acceptance criteria previously used for evaluation of the original plan. Plan-robustness-analyses, in order to determine the treatment plan that is most robust against variations in anatomy and patient setup, will help to minimize the effect of treatment uncertainties.

In chapter 6 we showed that an advanced 6-beam split-field scanned IMPT technique more adequately avoided density heterogeneities and therefore probably improves the plan robustness. This study investigated the clinical feasibility of such a split-field technique. The robustness of this approach to anatomical changes will be the subject of future work.

Analyses regarding the robustness of scanned IMPT treatment of head and neck cancer are scarce. None of the discussed ISPC studies (chapter 2) compared IMRT and IMPT plan robustness. Some studies that did consider proton plan robustness analyses are discussed below. It should be emphasized that the application of safety margins (expanding the clinical target volume) to deal with setup variations and organ motion, which is common in photon therapy treatment planning, is probably not the most suitable strategy for handling uncertainties in IMPT treatment planning [4,177]. Albertini et al. [4] showed that the application of safety margins around the PTV may help to improve plan robustness for IMPT without highly modulated dose distributions per field. However for plans with highly modulated dose distributions those safety margins only marginally improved plan robustness. Thus, more advanced tools are necessary.

The sensitivity of IMPT plans to range uncertainties, inter-fraction and intra-fraction motion has been discussed in three studies [3,113,114]. Lomax et al. [113] simulated range uncertainties by recalculating plans on a modified CT scan in which the CT values were changed by ± 3%. Robustness of Distal Edge Tracking (DET) IMPT plans were compared with the robustness of 3D-IMPT plans. (The 3D-IMPT planning strategy refers to the same strategy used for the IMPT plans discussed in this thesis.) 3D-IMPT planning evenly distributes the Bragg peaks in three dimensions over the target volume for each field. Subsequently, the TPS optimizes all Bragg peaks. In contrast, DET-IMPT planning only considers the Bragg peaks located at the distal edge of the target volume for each field in the optimization. DET-IMPT turned out to be relatively sensitive to the simulated CT errors, whereas 3D-IMPT was quite robust, unless strong internal dose gradients were present [113]. For the skull base case included in that study, compared with

the nominal plans (the dose distribution results based on the original planning CT scan), only small differences were found for the 3D-IMPT plans when analysing the dose in the CTV and corresponding DVH data, whereas for the DET plans (versus nominal plans) larger differences were found. More specifically, clear hot areas occurred for increased CT values (simulating an undershoot of the Bragg peak) within the CTV, resulting in a shift of the DVH curve towards higher doses by 5% (for decreased CT values, simulating an overshoot of the Bragg peak, cold areas [<95% prescribed dose] occurred within the CTV). In a subsequent study [114], inter-fraction and inter-field errors were simulated by shifting the CT data set and by shifting the dose distribution for each field direction, respectively. The 3D-IMPT plans were generally more robust than the DET-IMPT plans. For instance, for the skull base case, the worst case scenario for an inter-field motion error displayed an underdosage of the PTV only at its border with 3D-IMPT, whereas with DET-IMPT clear underdosage was also found within the CTV. However, again the presence of strong dose gradients in the fields reduced plan robustness.

In another study, Albertini et al. [3] used an anthropomorphic head and neck phantom (assuming a skull base tumour) to analyse plan accuracy in the presence of high heterogeneity, and plan robustness in case of range and spatial treatment uncertainties. Range errors were simulated by modification of the CT values by ± 3% while spatial errors were simulated by rotating the phantom by 3° along the cranial-caudal axis (simulating the effect of a residual patient setup error). Film measurements were compared with the calculated dose distributions. Both 3D-IMPT and DET-IMPT plans were analysed. For both range as well as spatial errors, 3D-IMPT plans were more robust than DET-IMPT plans and both plans were more robust to range errors than to the spatial errors. The outcomes of a recently developed model that predicts the delivery errors due to range errors [4] agreed with the measured errors, both for DET- and 3D-IMPT plans. As range error uncertainties are probably the main source of error [4], this tool will probably be very helpful in the evaluation of plan robustness.

The previously discussed studies indicate that strong dose gradients in the fields can adversely affect plan robustness, especially when they are located in a

very heterogeneous region (e.g. including soft tissue–air-bone transitions). Variations in patient alignment or anatomy will affect the dose distribution in these regions substantially more than in regions with less heterogeneity.

The plan-robustness-analyses tool as proposed by Albertini et al. [4] did not directly improve IMPT plan robustness in treatment planning. Therefore, the question arises as to whether there are tools that are able to increase plan robustness. Albertini et al. [4] indicated that the beam configuration significantly affects plan robustness. However, automatic beam angle optimization is not yet available in TPSs for proton therapy. Therefore, the selection of beam angles is currently based on utility tools provided by the TPSs (e.g. tools that indicate the density heterogeneity along the beam path) and on the experience of the treatment planner. However, tools that directly incorporate range uncertainties and setup errors in the plan optimization algorithm (implemented in the TPS) are currently under development [144,177,178], e.g. by using probabilistic planning methods that take into account various error scenarios during optimization. In such a probabilistic treatment plan optimization approach, the dose distribution directly depends on a set of random variables that simulate the uncertainty in the relevant parameter (e.g. the uncertainty in proton range), and hence, the corresponding effect of this variation on the dose distribution is taken into account directly during plan optimization. Studies investigating those tools all demonstrated that direct incorporation of these uncertainties into the optimization yields more robust treatment plans [144,177,178]. In general, the more robust plans avoided (1) high dose gradients in beam direction and the distal falloff of the Bragg peak to be directly in front of an OAR, and (2) high dose gradients in the lateral direction within a field. In order to improve treatment plan robustness, tools like these should be available in the TPSs used for both photon and proton therapy.

In the ISPC studies presented in this thesis, we used the same safety margins for proton and photon plans to deal with setup errors and organ motions. However, as previously mentioned, these safety margins might not be suitable for handling those uncertainties in scanned proton therapy. As demonstrated in a recent study, in case of protons, beam-specific PTV margins may improve plan robustness [137].

Moreover, Park et al. [137] demonstrated that proton plans based on beam-specific PTV margins resulted in better target coverage in presence of range uncertainties and setup errors. Unfortunately, this newly introduced method [137] is not yet applicable for multi-beam plans and/or IMPT plans. Hence, further research is necessary to investigate the optimal way to incorporate the beam-specific planning margins to improve treatment plan robustness for multibeam and/or IMPT proton therapy plans. Additionally, it should be taken into account that using beam-specific planning margin for plans consisting of multiple beams will result in the definition of multiple (beam-specific) PTVs. Hence, a new plan analyses method should be developed too, to be able to adequately analyse those new treatment plans based on beam-specific planning margins.

Taking into account the results of the studies discussed above, we can conclude that the ISPC results presented in this thesis may become different when incorporating those treatment plan robustness tools into the planning optimization. Moreover, in the ISPC study analyses, only planning CT scans were used and hence these analyses did not take into account/simulate the influence of setup errors and range uncertainties that may occur during the course of fractionated radiation treatment. Second, we used a uniform CTV-PTV margin (just like all the other ISPC studies as discussed in chapter 3) instead of beam-specific PTV margins. Consequently, implementation of (1) treatment plan robustness tools (directly incorporated into the plan optimization) and/or (2) beam-specific PTV margins to improve treatment plan robustness, may yield different results and may either lead to increased or decreased doses to the salivary glands and the SWOARs. The influence of these plan robustness improvement methods on the OAR-sparing capabilities of IMPT in HNC, as demonstrated in this thesis, will be investigated in further research.

Fractionation and beam arrangement

Cozzi et al. [41] performed an ISPC study that investigated the potential benefit of a full simultaneous integrated boost (SIB) technique in proton therapy compared with a conventional sequential (SEQ) boost technique and a mixed

SEQ/SIB technique. In the SIB technique, two different dose levels are irradiated simultaneously during the full treatment course. In the SEQ technique, during the first part of the treatment course the elective PTV (that also included the boost PTV) is irradiated up to its prescribed dose, while in the second part only the boost PTV is irradiated. The SEQ/SIB technique concerns an alternated scheme, starting off with a SEQ technique, irradiating the elective PTV, and continuing with a SIB technique resulting in an accelerated treatment for the boost PTV. The SEQ/SIB technique showed a dosimetric advantage in terms of target coverage while keeping similar OAR dose results. This technique appears promising given the logistic advantages (lower number of fractions) and the limited availability of proton therapy facilities. Moreover, fewer fractions mean a shorter treatment course that is more convenient for patients and allows for a higher throughput of patients. However, more studies on this subject are necessary.

Steneker et al. [170] showed in a previous study that target dose homogeneity and OAR-sparing were similar with 3-, 5- and 9-field IMPT. However, this study only considered the low dose PTV and attempted to spare only two OARs, i.e. the parotid glands and the spinal cord. In contrast to the results found by Steneker et al. [170], the results in this thesis showed that increasing the number of beams improved sparing of OARs when multiple OARs (salivary glands and SWOARs) and more complex targets are involved (chapter 7). This can be explained by the fact that increasing the amount of beams inherently increases the degrees of freedom regarding sparing of the multiple OARs (while simultaneously maintaining adequate target coverage). However, increasing the number of beams does not directly improve plan robustness, while this is of major importance regarding the clinical implementation of protons for head and neck cancer cases.

Fortunately, the recently developed plan-robustness-analyses tool [4] and the additional tools under development [137,144,177,178] will enable the improvement of treatment plan robustness by selecting the optimal number of beams and their geometry.

Interpretation of NTCP outcome results

The results of this thesis showed that, by applying IMPT, the dose to the submandibular glands, the parotid glands and the SWOARs can all be reduced. Existing NTCP models indicated that these dose reduction results translated into reduced probabilities of xerostomia, salivary flow dysfunction and physician-rated and patient-rated swallowing dysfunction. However, there are some issues that should be discussed with regard to the predicted NTCP reductions.

First, NTCP reductions are not only determined by a decrease in mean dose, but also by the absolute dose to the OAR as initially obtained with the reference technique and by the shape of the NTCP curve. When the initially obtained mean dose (reference technique) is already very high or very low, and therefore located near/in the plateau region of the NTCP curve (see Figure 1, chapter 3), a decrease in mean dose of for instance 10 Gy would yield a much smaller benefit as compared to the case in which the initial mean dose value is located near/in the steep part of the NTCP curve.

Second, the potential clinical benefit of the application of proton therapy in terms of NTCP varies among patients, which indicates that reporting only population-based mean differences in dose distribution between two radiation techniques will not be sufficient to explore the full potential or possible disadvantages of these techniques. Goitein stated that proton dose distributions are in almost all cases superior to photon dose distributions, and only with very few exceptions there is little clinical reason to argue against the use of protons for any patient or tumour site [70]. However, the results as presented in this thesis indicate that this statement requires some refining. Probably it is true that for each disease site there will be a certain percentage of patients that benefit from proton therapy, but the degree of this benefit will vary among individual patients.

Third, the NTCP predictions should be interpreted with some caution, for two major reasons. First, models that were developed in one or more institutes do not necessarily yield accurate predictions for patients treated in other institutes. Moreover, the NTCP predictions for proton treatment plans were calculated with models that were based on 3D-CRT and/or IMRT (photon) data. Since the dose distributions achieved with proton therapy differ from those obtained with photon

therapy, the NTCP predictions may not be accurate for proton therapy. Dose distributions achieved with 3D-CRT photons generally exist of relatively large volumes that receive a homogeneous dose (see Figure 1, chapter 4), including the targets and closely surrounding OARs. In contrast, IMRT dose distributions are much more inhomogeneous, especially just outside the target (border targets - OARs). Recently, Dijkema et al. [46] illustrated that NTCP curves for salivary flow, based on the mean parotid gland dose, are different for 3D-CRT and IMRT patients. More recently, Beetz et al. [11] showed that predictive models based on 3D-CRT data [10], for patient-rated xerostomia and sticky saliva, gave less accurate predictions for patients treated with IMRT. This illustrates that models based on a specific technique should be used with caution for patients treated with another technique and should in fact be validated again among patients treated with the new radiation delivery technique. Dijkema and Beetz illustrated the need for properly designed clinical studies to validate whether NTCP models based on photon treatment can be extrapolated to proton treatment. Fortunately, NTCP predictions for proton therapy may be relatively accurate when obtained with NTCP models based on IMRT treatment. The differences in the dose distributions between scanned IMPT and IMRT are smaller (Figure 1, chapter 4) than the differences between IMRT and 3D-CRT dose distributions. Therefore, we expect that the NTCP models based on IMRT data will better agree with the NTCP models based on IMPT data than with the models based on 3D-CRT data. Still, the fraction doses to the parotid glands are substantially reduced in IMPT compared with IMRT, which may yield even less complications with IMPT than currently predicted by the IMRT-based NTCP models.

In addition to dose distribution dependencies, variations in regional sensitivity to radiation of a specific organ can also influence the amount of complications. Animal studies have shown regional variations in sensitivity in both lung and parotid glands [98,101,102,110]. The parotid glands of rats were found to be much more sensitive in the cranial than in the caudal part and, additionally, irradiation of only the cranial part caused secondary damage to the unirradiated part [101,102]. Similar phenomena may occur in humans. As the dose distributions achieved in the

glands differ for different irradiation techniques, the salivary gland damage may therefore differ as well.

Conclusions

The results as presented in this thesis showed that IMPT has the potential to significantly reduce the risk on salivary gland and swallowing dysfunction and may subsequently improve quality of life for the majority of head and neck cancer patients, as compared with the currently used most advanced IMRT techniques.