



Clinical Monte Carlo versus Pencil Beam Treatment Planning in Nasopharyngeal Patients Receiving IMPT

Balu Krishna Sasidharan, MBBS, MD¹; Saif Aljabab, MD²; Jatinder Saini, PhD³; Tony Wong, PhD³; George Laramore, PhD, MD²; Jay Liao, MD²; Upendra Parvathaneni, MD²; Stephen R. Bowen, PhD⁴

¹Radiation Oncology, Christian Medical College, Vellore, Tamil Nadu, India

²Radiation Oncology, University of Washington/Seattle Cancer Care Alliance Proton Therapy Center, Seattle, WA, USA

³Seattle Cancer Care Alliance Proton Therapy Center, Seattle, WA, USA

⁴Radiation Oncology and Radiology, University of Washington School of Medicine, Seattle, WA, USA

Abstract

Purpose: Pencil beam (PB) analytical algorithms have been the standard of care for proton therapy dose calculations. The introduction of Monte Carlo (MC) algorithms may provide more robust and accurate planning and can improve therapeutic benefit. We conducted a dosimetric analysis to quantify the differences between MC and PB algorithms in the clinical setting of dose-painted nasopharyngeal cancer intensity-modulated proton radiotherapy.

Patients and Methods: Plans of 14 patients treated with PB analytical algorithm optimized and calculated (PBPB) were retrospectively analyzed. The PBPB plans were recalculated using MC to generate PBMC plans and, finally, reoptimized and recalculated with MC to generate MCMC plans. The plans were compared across several dosimetric endpoints and correlated with documented toxicity. Robustness of the planning scenarios (PBPB, PBMC, MCMC) in the presence of setup and range uncertainties was compared.

Results: A median decrease of up to 5 Gy ($P < .05$) was observed in coverage of planning target volume high-risk, intermediate-risk, and low-risk volumes when PB plans were recalculated using the MC algorithm. This loss in coverage was regained by reoptimizing with MC, albeit with a slightly higher dose to normal tissues but within the standard tolerance limits. The robustness of both PB and MC plans remained similar in the presence of setup and range uncertainties. The MC-calculated mean dose to the oral avoidance structure, along with changes in global maximum dose between PB and MC dosimetry, may be associated with acute toxicity-related events.

Conclusion: Retrospective analyses of plan dosimetry quantified a loss of coverage with PB that could be recovered under MC optimization. MC optimization should be performed for the complex dosimetry in patients with nasopharyngeal carcinoma before plan acceptance and should also be used in correlative studies of proton dosimetry with clinical endpoints.

Submitted 19 Oct 2018
Accepted 07 March 2019
Published 18 April 2019

Corresponding author:

Stephen R. Bowen
Departments of Radiation
Oncology and Radiology
(joint)
University of Washington
School of Medicine
1959 NE Pacific St Box
356043
Seattle, WA 98195, USA
Phone: +1 (206) 543-6559
srbowen@uw.edu

Original Article

DOI
10.14338/IJPT-18-00039.1

© Copyright
2019 The Author(s)

Distributed under
Creative Commons CC-BY

OPEN ACCESS

<http://theijpt.org>

Introduction

With the advent of advanced radiation treatment facilities across the world, proton beam radiotherapy is increasingly being used for radical intent therapy of head and neck cancers. It offers the benefit of lower integral body dose in the absence of an exit beam compared with conventional photon beam radiotherapy [1, 2]. Head-and-neck radiotherapy presents a unique treatment planning challenge with multiple prescribed dose levels and several critical organs close to targets. The presence of bone soft-tissue interfaces, oral cavity, and pharyngeal airway can also add to the inconsistency with pencil beam (PB) analytical dose calculation in head-and-neck proton beam radiotherapy [3]. Hence, the accuracy of dose delivery and dose calculation is important in patient management. Changes in critical organ dosimetry may influence risks of acute and chronic toxicity. Similarly, failure to provide optimal target coverage may prove detrimental to tumor control.

The PB analytical dose engines have been the standard for proton radiotherapy as it offers a fast optimization process and efficient computation. However, its accuracy in heterogeneous phantom tissues is compromised compared with recent commercial Monte Carlo (MC) dose engines [4, 5]. Furthermore PB treatment planning is often used in lieu of MC treatment planning to gain speed and efficiency and to achieve clinical dose constraints. Maes et al. [6] observed deficiencies in planned dosimetry for lung cancer proton radiotherapy plans due to target heterogeneity and recommended integrating the MC algorithm into clinical planning workflows. This dosimetric comparison between MC- and PB-optimized plans has not been quantified in the clinical setting of head and neck cancer. The impact of and tradeoffs between PB and MC treatment-planning approaches may provide useful decision support to head-and-neck radiation oncologists.

The purpose of this investigation was to evaluate and quantify dosimetric differences between PB and MC treatment planning approaches in a retrospective series of patients with nasopharyngeal cancer. We also correlated these changes in dosimetry with clinical toxicity endpoints and evaluated the robustness of planning approaches under range and patient setup uncertainties.

Patients and Methods

Patient Cohort

The treatment-planning data from 14 patients diagnosed with nasopharyngeal cancer who received definitive proton pencil beam scanning (PBS) radiotherapy were reviewed and analyzed. These patients were enrolled into an Institutional Review Board–approved trial registry at the University of Washington. For this cohort of patients, the gross tumor volumes ranged from 23 to 63 cm³ with a median volume of 39 cm³ for the primary tumor. The nodal GTV ranged between 2 and 20 cm³ with a median volume of 11.5 cm³. All patients successfully completed therapy with a minimum of 3 months' follow-up. Toxicity data (CTCAE V 4.0) were collected from patient charts for correlation with the treatment-planning dosimetry. Patient characteristics are summarized in Table 1.

Patient Simulation

Patients were simulated using the Optima CT580 (GE Healthcare, Waukesha, Wisconsin) and scanned headfirst in supine position, immobilized on a BoS table (Qfix, Avondale, Pennsylvania) with thermoplastic mask, custom moldcare cushion, and custom oral stent. Helical planning computed tomography (CT) scans were acquired and reconstructed on a 65-cm transverse field-of-view (FOV) with 2.5 mm slice thickness. Contrast-enhanced helical CT scans over 35 cm FOV and metal artifact-reduced CT reconstructions over 50 cm FOV were also generated and fused for treatment planning. Contouring of targets and organs at risk (OARs) was completed in MIM 6.6 (MIM Software, Cleveland, Ohio).

Treatment Planning and Optimization Techniques

Analytical PB-based treatment planning was performed using PB dose version 4.1, while MC-based treatment planning was performed using the MC Dose Version 4.0 algorithm in the RayStation 6.0 SP1 (RS6) (RaySearch Laboratories, Stockholm, Sweden). The planning target volume (PTV) margin expansion was not beam-specific and nominally consisted of an isotropic expansion of 3 to 5 mm. Clinician-defined margins accounted for nasal cavity filling changes in certain cases. Plans were routinely optimized and subsequently evaluated for dosimetric differences with a nasal cavity override turned on versus off, in order to simulate the worst-case scenario in the presence versus absence of nasal cavity obstruction. In general, cases that necessitated plan adaptation were the result of significant decreases in nasal cavity filling during treatment.

Planning configurations varied from 2- to 5 beams but primarily consisted of 3 beams arranged either in Y (PA, anterior obliques) or inverted-Y (AP, posterior obliques) orientations. In cases of complex bilateral target geometry, single-field optimization with split target junction matching was utilized to enhance sparing of critical structures. Due to the shallow proximal extent of nasopharyngeal tumors, range shifters were used in every case, and the air gap between range shifter and patient surface was minimized while avoiding collisions. Objectives for target and normal tissue dosimetry were matched to the original clinically approved plan in each case. Optimization consisted of a hybrid set of PTV and/or robust clinical target volume (CTV) objectives, which varied for each clinical case. Robustness settings for CTV objectives included range ($\pm 3\%$ to 3.5%) and setup (3 to 5 mm) perturbations. The reoptimization with MC algorithm used the same robustness evaluation criteria as the clinically approved base plan, PBPB. Dose painting prescriptions for definitive proton beam therapy regimens included 66 to 70 Gy(RBE) to the PTV high risk (PTV HR), 60 to 63 Gy(RBE) to the PTV intermediate risk (PTV IR), and 54 Gy(RBE) to the PTV low risk (PTV LR) in 33 fractions. The dose calculation grid during optimization and final computation was fixed at an isotropic 2 mm for all plans, per our institutional standard for head-and-neck disease sites. The critical organ doses were kept below standard-of-care guidelines. All doses were calculated and reported with an RBE of 1.1, per current clinical practice.

Clinically approved PB optimized plans were recalculated using RS6 MC algorithm to obtain PBMC plans. For MC treatment planning, under consistent optimization objectives relative to the clinical PB plan, a sampling history of 10 000 ions/spot was imposed along with a sufficient number of ions to yield 0.5% statistical uncertainty in the final calculated dose (MCMC). The robustness criteria used for baseline PB optimization were retained for MC optimization. The PBPB and MCMC plans were normalized to achieve prescribed dose coverage to 95% of the planning target volumes. The PBMC plans were not normalized as they were recalculated using same beam spot size and spot weight as the original PBPB plan. The MC reoptimization of PB plans used a minimum of 100 iterations. The statistical uncertainty was fixed at 0.5% for all MC calculations (PBMC and MCMC).

Dosimetric and Statistical Analysis

The differences in target coverage and normal tissue dose were tabulated and compared between the PBPB, PBMC and MCMC plans. The following plan metrics defined in RS6 were evaluated: PTV HR Conformity Index (CI) and Homogeneity Index (HI), PTV HR D_{95} , PTV IR D_{95} , PTV LR D_{95} , Brainstem $D_{0.03\text{ cm}^3}$, Spinal cord $D_{0.03\text{ cm}^3}$, and total parotid D_{Mean} .

The CI was calculated using the equation

$$(1) \quad \text{Conformity Index} = (\text{PTV HR volume receiving } 70\text{Gy}) / (\text{PTV HR Volume})$$

The HI was calculated using the equation

$$(2) \quad \text{Homogeneity index} = D_{95} / D_5 \text{ of the target}$$

Pairwise differences in planned dose and the dose-volume parameters among the 3 groups of plans (PBPB, PBMC, MCMC) were evaluated using nonparametric Wilcoxon signed rank tests.

To assess treatment plan robustness to setup and range uncertainties, the perturbed dose distributions for the 3 planning methods were generated and compared. Perturbation scenarios consisted of under-/over-ranging beams by scaling the CT density $\pm 3\%$ and shifting beam isocenters ± 3 mm in the medial/lateral (x), anterior/posterior (y), and superior/inferior (z) directions. Two representative worst-case dose perturbations scenarios described by (+3%, x + 3 mm, y + 3 mm, z + 3 mm) and (-3%, x - 3 mm, y - 3 mm, z - 3 mm) were compared and analyzed. Relative changes in dosimetry between the worst-case perturbed plan and nominal plan were analyzed to assess the variability in robustness across PBMC and MCMC plans. Pairwise differences between PBMC and MCMC perturbed dose deviations from nominal plans were evaluated with nonparametric Wilcoxon signed rank tests.

Dosimetric changes in OARs between the original clinical PB plan and the PBMC recalculated plan were correlated to toxicity. The OAR dose parameters included mean doses to oral avoidance and laryngo-pharynx avoidance structures, as well as global dose maximum. Toxicity scoring was dichotomized based on incidence of acute grade 3 or higher radiation dermatitis, acute grade 3 or higher radiation mucositis, completion of prescribed chemotherapy and requirement for interventions (emergency department, intensive care unit, inpatient care, feeding procedure, unscheduled intravenous fluids, etc). Differences in OAR dosimetry between patients who experienced toxicity and those who did not were evaluated with

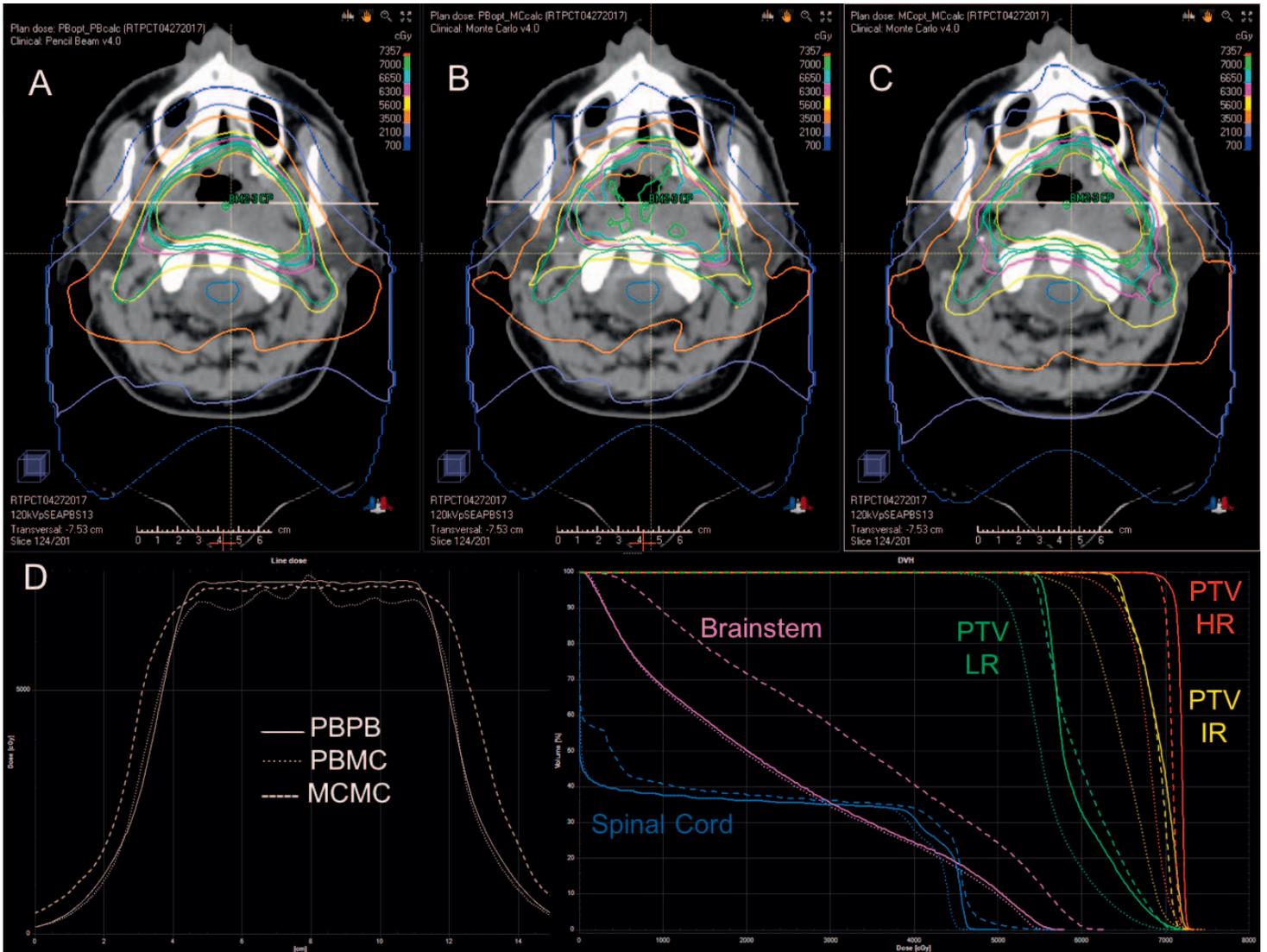


Figure. Representative case of 1 patient with nasopharynx cancer planned for proton PB scanning therapy using 3 techniques: (A) clinically approved analytic PB optimization and calculation PBPB. (B) PB optimization and MC calculation PBMC (C) MC optimization and calculation MCMC. The upper row shows the dose distributions (rainbow isodose lines) for each planning technique overlaid on the planning computed tomography and target volume (high-risk planned target volume, red). (D) The bottom row shows the line dose profiles (transverse-white) for each of the 3 techniques (solid line: PBPB, dotted line: PBMC, dashed line: MCMC) and the corresponding dose-volume histograms. Abbreviations: MC, Monte Carlo; PB, pencil beam.

nonparametric Mann-Whitney tests. For nonparametric test results, $P \leq .05$ were considered statistically significant. All statistical analyses were performed using OriginPro 9.1 (OriginLab Corporation, Northampton, Massachusetts).

Results

The Figure illustrates isodose distributions overlaid on a patient planning CT under three planned dose conditions: (A) PBPB, (B) PBMC, and (C) MCMC. The line dose profile (D) in craniocaudal and transverse planes reveals a consistent deviation in coverage with PBMC compared with the baseline PBPB plan. This deviation in dosimetry is recovered with the MCMC plan, both on the line profile of Figure D and the dose-volume histogram on Figure E.

Table 2 lists target and critical structure dose differences between the PBPB, PBMC, and MCMC treatment plans. The PTV volumes of the patients ranged from 73 to 179 cm³ with a median of 144 cm³ for high risk, 131 to 391 cm³ with median of 233 cm³ for intermediate risk and 237 to 514 cm³ with median of 337 cm³ for low-risk regions. For PBPB plans, the PTV HR D95 ranged from 67 to 70 Gy with a median of 70 Gy, PTV IR D95 ranged between 62 and 64 Gy with a median of 64 Gy and PTV

Table 1. Patient characteristics.

Characteristic	
Age	Median: 58 years (IQR, 46-63)
Sex	
Men	11
Women	3
Ethnicity	
Asian	7%
Caucasian	5%
Karnofsky performance status	All 90 at start of radiation therapy
Gross tumor volume	
Primary	39.34 cm ³ (IQR, 22.73-62.73)
Nodal	11.47 cm ³ (IQR, 4.08-26.04)
Prescription	66-70 Gy to PTV HR in 33 fractions 60-63 Gy to PTV IR in 33 fractions 54 Gy to PTV LR in 33 fractions
Concurrent Chemotherapy	
Yes	14 (100 %)
No	0 (0%)
Overall radiation therapy treatment duration	44 days (IQR,44-44.75)

Abbreviations: IQR, interquartile range; PTV HR, planning treatment volume high risk; PTV IR, planning treatment volume intermediate risk; PTV LR, planning treatment volume low risk.

LR D95 had a median of 55 Gy within a narrow range of 54 to 55 Gy. Under PBMC plans, the PTV HR D95 ranged from 63 to 66 Gy with a median of 65 Gy, and PTV IR D95 ranged from 56 to 60 Gy with a median of 58 Gy. Both PTV HR and PTV IR D95 on the PBMC plans showed a statistically significant drop in target coverage relative to PBPB plans ($P < .005$). The median percentage drop in coverage for CTV primary was 4.7% (interquartile range [IQR], 3.4–4.9) and for CTV neck was 4.3% (IQR, 3.0–6.1). There also appear to be significant cold spots of range 4 to 6 Gy within high risk CTV with MC recalculation, which would be considered clinically significant.

The PBMC plan had a 5.8% (IQR, 5.2%–7.0%) drop in coverage relative to prescription for PTV HR compared with the PBPB plan. The drop in coverage was 6.1% (IQR, 4.8%–9.9%) and 7.3% (IQR, 5.4%–10.2%) for PTV IR and PTV LR, respectively. Interestingly, when we compared the drop in coverage with PBMC from baseline PBPB plans with MC recalculation, the median percentage drop in coverage was 12.2, 11.7, and 9.6 for 5 beam plans, and 5.5, 6.0, and 7.1% for 3 beam plans for PTV HR, IR, and L,R respectively.

Table 2. PTV HR, PTV IR, and PRV LR D95; PTV HR homogeneity index; PTV HR Conformity Index; brainstem D0.03 cm³; spinal cord D0.03 cm³; total parotid D_{mean} for 14 previously treated patients with nasopharynx cancer calculated with PB and MC. Bolded p-values indicate statistical significance.

Dosimetric parameters	PBPB		PBMC		Wilcoxon signed rank PBPB vs PBMC, P value	MCMC		Wilcoxon signed rank PBPB vs MCMC, P value
	Median	IQR	Median	IQR		Median	IQR	
PTV HR D95	69.75	67–70	65.33	63–66	.002	69.37	67–69	.22
PTV IR D95	63.44	62–64	58.11	56–60	.003	63.11	61–64	.09
PTV LR D95	54.51	54–55	50.32	49–51	.006	54.17	52–55	.11
PTV HR CI	0.85	0.76–0.88	-	-		0.84	0.74–0.90	.84
PTV HR HI	0.96	0.91–0.97	0.91	0.90–0.92	.002	0.96	0.93–0.97	.57
Brainstem D0.03 cm ³	57.81	49–61	57.47	53–63	.391	60.14	51–62	.3
Spinal cord D0.03 cm ³	41.26	34–46	39.96	35–47	.116	39.48	31–49	.71
Total Parotid D _{mean}	33.72	25–40	31.53	24–38	.005	34.73	26–41	.002
Global Dose max	73.24	72–74	73.13	71–77	.855	73.52	73–74	1.00

Abbreviations: PTV HR, planning treatment volume high risk; PTV IR, planning treatment volume intermediate risk; PTV LR, planning treatment volume low risk; HI, homogeneity index; CI, conformity index; PB, pencil beam; MC, Monte Carlo; PBPB, pencil beam optimized; PBMC, Monte Carlo recalculated; MCMC, Monte Carlo optimized; IQR, interquartile range.

Table 3. Change in PTV HR D95, PTV IR D95, PTV LR D95, and brainstem maximum dose under the following perturbed scenarios relative to the nominal plan: (+3% density/under-range, x + 3 mm, y + 3 mm, z + 3 mm), (−3% density/over-range, x − 3 mm, y − 3 mm, z − 3 mm).

Worst-case median change in dose, Δ3%/3mm	PBMC		MCMC		Wilcoxon signed rank, P
	Median	IQR	Median	IQR	
Δ PTV HR D95	−2.9	[−3.9, −2.2]	−2.9	[−3.7, −2.3]	.6
Δ PTV IR D95	−3.4	[−3.9, −2.7]	−3.7	[−4.7, −2.9]	.9
Δ PTV LR D95	−2.7	[−3.3, −2.0]	−3.0	[−3.4, −1.8]	.9
Δ brainstem maximum	3.4	[+2.3, +9.0]	3.5	[+1.8, +5.9]	.2

Abbreviations: PTV HR, planning treatment volume high risk; PTV IR, planning treatment volume intermediate risk; PTV LR, planning treatment volume low risk; PBMC, Monte Carlo: recalculated; MCMC, Monte Carlo: optimized; IQR, interquartile range.

The HI for PTV HR was also significantly lower on the PBMC plan ($P < .005$). For MCMC plans, PTV D95 metrics, PTV CI, and PTV HI were concordant to the values achieved in PBPB plans, with no statistically significant differences ($P > .09$). The total parotid Dmean was lower in the PBMC plan with respect to the PBPB plan ($P = .005$) and higher in the MCMC plans ($P = .002$). The PTV CI, Brainstem D0.03 cm³, and spinal cord D0.03 cm³ were not statistically different among the dose calculation and optimization combinations. The median Global D_{max} was consistent among all plans.

Table 3 shows delta values representing the percentage change in PTV HR D95, PTV IR D95, PTV LRD95, and maximum dose to the brainstem (0.03 cm³) in PBMC and MCMC treatment plans following range and isocentric setup perturbations: (+3% density/under-range, x + 3 mm, y + 3 mm, z + 3 mm), (−3% density/over-range, x − 3 mm, y − 3 mm, z − 3 mm). All changes were calculated relative to unperturbed nominal plans.

The median decrease in the PTV HR D95 coverage was 2.9% for both methods, in PTV IR D95 it was 3.4% for PBMC and 3.7% for MCMC, and in PTV LR D95 it was 2.7% and 3% for the PBMC and MCMC plans, respectively. The worst-case change in brainstem maximum dose was a median increase of 3.4% in PBPB and 3.5% in the MCMC plan. No changes in dosimetric parameters were statistically significant, suggesting that robustness to range and setup uncertainties was similar between PM- and MC-based treatment planning.

Table 4 compares the median change in dose when PBPB plan was recalculated to give the PBMC plan. Four of 14 patients (29%) were unable to complete the prescribed course of chemotherapy. There was an increase in global D_{max} by 2.8% for the suboptimal chemotherapy group compared with a drop of 1.2% for those who completed chemotherapy, but this numerical difference was not statistically significant ($P = 0.28$). Similarly, correlating the mean doses of the oral avoidance structure and the laryngeal avoidance structures did not show any significant difference. A total of 10 patients developed significant

Table 4. Correlating incidence of acute events to calculated change in plan dosimetry of approved PBPB plan on recalculated plans with MC algorithm (PBMC).

Event (%)	Dose parameter	Median PBMC dose, cGy RBE					Median change in dose with PBMC, %				
		No	IQR	Yes	IQR	P*	No	IQR	Yes	IQR	P*
Unable to complete prescribed chemotherapy (28)	Global D _{max}	7242	7119–7557	7708	7434–7944	.10	−1.2	−2.7 to +1.9	+2.8	−1.3 to +4.8	.28
	Oral avoidance D _{mean}	2226	1525–3084	1139	998–1797	.02	0.0	−2.0 to +1.8	−1.8	−2.4 to +0.06	.43
	Larynx D _{mean}	2579	1972–3226	1915	1330–3321	.53	−0.4	−1.7 to +2.4	−2.4	−4.2 to +1.3	.62
More than grade 2 mucositis (71)	Global D _{max}	7150	7104–7640	7415	7255–7893	.18	−1.9	−3.0 to +3.6	−0.1	−2.5 to +4.3	.44
	Oral avoidance D _{mean}	1892	1149–3039	1835	1292–2935	1.00	−2.6	−4.6 to −1.5	−0.2	−2.3 to +1.8	.18
	Larynx D _{mean}	2441	1310–2955	2469	1250–3364	.73	−0.4	−1.5 to +2.2	−1.3	−3.2 to +2.4	.72
More than grade 2 dermatitis (50)	Global D _{max}	7313	7173–7796	7373	7097–7994	1.00	−2.4	−3.0 to −0.5	+2.0	−1.2 to +6.4	.09
Unscheduled events (42)	Global D _{max}	7290	7149–7711	7496	7125–8186	.57	−0.8	−2.2 to +4.0	−0.2	−3.0 to +7.0	.84

Abbreviations: PBMC, Monte Carlo recalculated; IQR, interquartile range.

*Mann-Whitney test.

mucositis (71%). There was a greater numerical drop of global D_{\max} and oral avoidance structure mean dose in the group who did not experience significant mucositis. Eight patients (57%) experienced unscheduled events, such as unplanned inpatient care, feeding intervention, unscheduled intravenous fluid administration, emergency department visits. A numerically lower dose in the recalculated plans was observed but did not reach statistical significance.

Discussion

It has been well demonstrated that MC dose calculation algorithms are a step toward improved proton beam radiotherapy dosimetric accuracy in the presence of tissue inhomogeneities. However, analytical PB dose computation engines remain popular due to fast optimization and widespread commercial availability in clinical treatment planning systems. Air gap influences differences between PB and MC dosimetry at the surface and depths <5 cm, due to how the nuclear halo and lateral scatter of secondary protons are modeled in the analytical dose engine. In general, analytical PB algorithms underestimate dose at shallower depths for the same spot position and intensity distribution. The effect scales with air gap and becomes significant with air gaps >15 cm. Beam central axis obliquity relative to the patient surface also influences dosimetric differences between algorithms at shallow depths. This phenomenon was studied at great length in our prior study [5, 7]. At depths >5 cm, dose differences are primarily driven by scatter through tissue inhomogeneities.

Yepes et al [8] conducted the largest comparison (125 patients) of MC dosimetry against standard treatment planning system analytic PB algorithms and reported absolute dose differences as high as 10 Gy in head-and-neck disease sites. However, they employed a nonclinical MC planning algorithm. Our study is the first to report on MC planning using a Food and Drug Administration approved and commercially available treatment planning system for head-and-neck cancers. Furthermore, this study looked at the influence of dose calculation algorithms in dose painting plans defined by multiple nested target regions (PTV HR, PTV IR, PTV LR) in the uniquely challenging nasopharynx disease site. We also have shown that the target dose coverage lost when recalculating an analytical PB-optimized plan with MC, representing a more accurate estimate of plan dosimetry can be regained through MC optimization. Hence, we highlight the need for clinically adopting MC planning in sites like the nasopharynx with complex dosimetry. Our results suggest overestimation of target dose coverage by analytical PB dose engines. Based on these findings, we advocate for the adoption of MC in routine clinical treatment planning and the potential for correlating dose differences between dose algorithms with clinical endpoints.

The median high-risk target volume in this cohort was 144 cm^3 , which is closer to the head-and-neck cohort reported by Yepes et al. (178 cm^3) [8]. Across 3-level dose painted plans we consistently observed a drop of 4 to 5 Gy in coverage of all targets between PB and MC dosimetry. This difference remained statistically significant for all dose levels. The loss of target coverage was regained in the MC optimized plans with all plans being able to achieve similar or better dose coverage compared with the baseline PBPB plans. It was noted that the drop in target coverage might be higher for greater beam numbers when MC recalculation of PB optimized plans is evaluated.

This increase in target dose coverage with MC plans was accompanied by a similar increase in normal tissue dosimetry with a sample median increase of 3 Gy in the brainstem maximum dose and median increase of 3 Gy in mean parotid doses. Despite an increase in normal tissue doses, all plans met clinical goals for standard-of-care head-and-neck plans. The global D_{\max} in MC-optimized plans remained statistically similar to PB-optimized plans despite numerical differences in parotid dosimetry.

The secondary analyses explored whether improved target coverage remains robust under the influence of range and setup uncertainties. We also tested whether higher perturbed dose to normal tissue adversely affected clinical planning goals and acceptance criteria. Our results revealed that variations in dose parameters under range/setup perturbation remain similar for target and critical structures irrespective of the optimization process. A study by Schuemann et al. [9] reported that the MC algorithms produced more accurate dosimetry in complex head-and-neck geometries and reduced uncertainties due to range differences and tissue heterogeneity. We similarly observed in our cohort that MC optimized plans were robust in the presence of range and setup uncertainties.

Dosimetric changes between the PB optimized plan and the MC recalculated plan were correlated with incidence of toxicity based on CTCAE V 4.0. A trend toward lower global maximum doses was observed in MC recalculated plans for patients with lower reported toxicity incidence, though this trend did not reach statistical significance. There have been few reported series of patients with nasopharyngeal carcinoma treated with proton therapy [10–12]. Most series report local control ranging from 75% to 100% and survival ranging from 70% to 90%. They also report toxicity and breaks during therapy. From our small patient sample, oral avoidance D_{mean} from MC dosimetry, along with global D_{\max} changes between PB and MC dosimetry,

may correlate with toxicity endpoints and should be explored further in future studies. With improved dosimetric accuracy from MC-based treatment planning, better estimates of toxicity risk factors during and after proton therapy may be identified.

Beyond increased accuracy of physical dose with MC, studies are also seeking to account for spatially variant linear energy transfer and relative biological effectiveness of proton treatment plans when correlating dosimetry with toxicity. These approaches require implementation of radiobiological models capable of predicting the location of toxicity [7, 13]. One key limitation of this work is that we have compared only the fixed relative biological effectiveness dose changes between algorithms. Due to the limited sample size, we could not definitively correlate toxicity incidence with plan dosimetry. A larger multi-institutional cohort may be required to address this question. With more patients, we plan to conduct a comprehensive study on proton MC dosimetry, extracted from PBMC and MCMC plans, for predicting toxicity in nasopharyngeal cancer patients.

Conclusion

Retrospective analyses of nasopharynx treatment plans showed statistically inferior dosimetric endpoints, affecting both target coverage and OAR doses, when using the analytical PB algorithms for treatment planning. The impact on dosimetric endpoints could be reversed through MC-based treatment planning. An MC optimization should be performed to characterize complex dosimetry in patients with nasopharyngeal carcinoma before plan acceptance and should also be used in correlative studies of proton dosimetry with clinical endpoints. Clinical integration of MC planning may lead to improved treatment tolerance and outcomes in patients with head-and-neck cancer.

ADDITIONAL INFORMATION AND DECLARATIONS

Conflicts of Interest: The authors have no conflicts of interest to disclose.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

1. Blanchard P, Gunn GB, Lin A, Foote RL, Lee NY, Frank SJ. Proton therapy for head and neck cancers. *Semin Radiat Oncol.* 2018;28:53–63.
2. Leeman JE, Romesser PB, Zhou Y, McBride S, Riaz N, Sherman E, Cohen MA, Cahlon O, Lee N. Proton therapy for head and neck cancer: expanding the therapeutic window. *Lancet Oncol.* 2017;18:e254–65.
3. Yamashita T, Akagi T, Aso T, Kimura A, Sasaki T. Effect of inhomogeneity in a patient's body on the accuracy of the pencil beam algorithm in comparison to Monte Carlo. *Phys Med Biol.* 2012;57:7673.
4. Saini J, Maes D, Bowen SR, Egan AJ, Bloch C, Wong TP. A comparison between pencil beam and Monte Carlo algorithms against film measurements in an anthropomorphic phantom for proton spot scanning. *Int J Radiat Oncol Biol Phys.* 2017;99:E717–8.
5. Saini J, Maes D, Egan A, Bowen SR, St James S, Janson M, Wong T, Bloch C. Dosimetric evaluation of a commercial proton spot scanning Monte-Carlo dose algorithm: comparisons against measurements and simulations. *Phys Med Biol.* 2017;62:7659–81.
6. Maes D, Bowen SR, Fung A, Saini J, Bloch C, Egan A, Zeng J, Rengan R, Wong TP. Dose comparison between proton pencil beam and Monte Carlo dose calculation algorithm in lung cancer patients. *Int J Radiat Oncol Biol Phys.* 2017;99:E694.
7. Maes D, Saini J, Zeng J, Rengan R, Wong T, Bowen SR. Advanced proton beam dosimetry part II: Monte Carlo vs. pencil beam-based planning for lung cancer. *Transl Lung Cancer Res.* 2018;7:114–21.
8. Yepes P, Adair A, Grosshans D, Mirkovic D, Poenisch F, Titt U, Wang Q, Mohan R. Comparison of Monte Carlo and analytical dose computations for intensity modulated proton therapy. *Phys Med Biol.* 2018;63:045003.
9. Schuemann J, Shin J, Perl J, Grassberger C, Verburg J, Faddegon B, Paganetti H. SU-E-T-500: pencil-beam versus Monte Carlo based dose calculation for proton therapy patients with complex geometries. Clinical use of the TOPAS Monte Carlo system. *Med Phys.* 2012;39:3820.

10. Chan AW, Liebsch LJ, Deschler DG, Adams JA, Vrishali LV, McIntyre JF, Pommier P, Fabian RL, Busse PM. Proton radiotherapy for T4 nasopharyngeal carcinoma. *J Clin Oncol*. 2004;22:5574.
11. Holliday EB, Garden AS, Rosenthal DI, Fuller CD, Morrison WH, Gunn GB, Phan J, Beadle BM, Zhu XR, Zhang X, Hanna E, Glisson BS, Hutcheson KA, El-Naggar AK, Hong J-H, Hung T-M, Uzel EK, Lewis G, Frank SJ. Proton therapy reduces treatment-related toxicities for patients with nasopharyngeal cancer: a case-match control study of intensity-modulated proton therapy and intensity-modulated photon therapy. *Int J Part Ther*. 2015;2:19–28.
12. Lewis GD, Holliday EB, Kocak-Uzel E, Hernandez M, Garden AS, Rosenthal DI, Frank SJ. Intensity-modulated proton therapy for nasopharyngeal carcinoma: decreased radiation dose to normal structures and encouraging clinical outcomes. *Head Neck*. 2016;38(suppl 1):E1886–95.
13. Fossum CC, Beltran CJ, Whitaker TJ, Ma DJ, Foote RL. Biological model for predicting toxicity in head and neck cancer patients receiving proton therapy. *Int J Part Ther*. 2017;4:18–25.