MONTE CARLO–BASED DOSIMETRY OF HEAD-AND-NECK PATIENTS TREATED WITH SIB-IMRT

NIRMAL SAKTHI, M.S., R.T.T., PAUL KEALL, PH.D., IVAYLO MIHAYLOV, PH.D., QIUWEN WU, PH.D., YAN WU, M.S., JEFFREY F. WILLIAMSON, PH.D., RUPERT SCHMIDT-ULLRICH, M.D.,† AND JEFFREY V. SIEBERS, PH.D.

Department of Radiation Oncology, Virginia Commonwealth University, Richmond, VA

Purpose: To evaluate the accuracy of previously reported superposition/convolution (SC) dosimetric results by comparing with Monte Carlo (MC) dose calculations for head-and-neck intensity-modulated radiation therapy (IMRT) patients treated with the simultaneous integrated boost technique.

Methods and Materials: Thirty-one plans from 24 patients previously treated on a phase I/II head-and-neck squamous cell carcinoma simultaneous integrated boost IMRT protocol were used. Clinical dose distributions, computed with an SC algorithm, were recomputed using an EGS4-based MC algorithm. Phantom-based dosimetry quantified the fluence prediction accuracy of each algorithm. Dose–volume indices were used to compare patient dose distributions.

Results and Discussion: The MC algorithm predicts flat-phantom measurements better than the SC algorithm. Average patient dose indices agreed within 2.5% of the local dose for targets; 5.0% for parotids; and 1.9% for cord and brainstem. However, only 1 of 31 plans agreed within 3% for all indices; 4 of 31 agreed within 5%. In terms of the prescription dose, 4 of 31 plans agreed within 3% for all indices, whereas 28 of 31 agreed within 5%.

Conclusions: Average SC-computed doses agreed with MC results in the patient geometry; however deviations >5% were common. The fluence modulation prediction is likely the major source of the dose discrepancy. The observed dose deviations can impact dose escalation protocols, because they would result in shifting patients to higher dose levels. © 2006 Elsevier Inc.

Monte Carlo, Simultaneous integrated boost, IMRT, Head-and-neck cancer, Dynamic multileaf collimators, Radiotherapy.

INTRODUCTION

Intensity-modulated radiation therapy (IMRT) has taken on a significant role in radiotherapy because of its ability to deliver radiation conformally around structures in a highly heterogeneous medium. Head-and-neck (HN) cancer patients are well suited to benefit from IMRT’s conformality, because the tumor volumes are often irregularly shaped and are directly adjacent to many critical structures. Studies have shown that compared with 3D conformal planning, IMRT is not only more conformal, but it results in a better dose homogeneity and a sharper dose gradient at the boundaries (1–6). Furthermore, IMRT planning techniques such as simultaneous integrated boost (SIB) can increase the conformality of treatment, shorten the overall treatment course, and limit the growth of aggressive diseases such as HN squamous cell carcinoma (6–11).

Several studies document the dosimetric benefits of IMRT combined with SIB techniques (12, 13). Previously, our institution reported on the dosimetric analysis of a group of HN patients treated with SIB-IMRT (11, 13). The dosimetric calculations in that report were based upon superposition/convolution (SC) algorithms with the intensity modulation incorporated into the dose calculation via a fluence transmission matrix. This method of including intensity modulation is similar to that implemented in commercial treatment planning systems (TPS) (14–17), although, instead of SC dose computation algorithms, pencil beam–based dose computation algorithms are more commonly used (18–20). Although beam-by-beam film-based quality assurance in a homogeneous phantom agreed with predictions using commonly accepted criteria in our previous study, discrepancies between the planned and delivered dose distributions were observed. However, the clinical impacts of these dose differences were not evaluated.

Reprint requests to: Jeffrey V. Siebers, Ph.D., Department of Radiation Oncology, Virginia Commonwealth University, 401 College Street, Richmond, VA 23298. Tel: (804) 628-7771; Fax: (804) 828-6042; E-mail: jsiebers@vcu.edu

†Rupert Schmidt-Ullrich, M.D., passed away in December 2004.

This work was supported by Grant CA98524 from the National Cancer Institute.

Acknowledgment—The authors would like to thank Devon Murphy-Stein for her meticulous editing of this manuscript.

Received Dec 7, 2004, and in revised form Aug 29, 2005. Accepted for publication Sept 4, 2005.
Intensity-modulated radiation therapy treatment fields, and SIB-IMRT treatment fields in particular, often have large intensity fluctuations that result in complex multileaf collimator (MLC) patterns and present challenges to dose calculation algorithms, because of the effects of radiation transmitted through and scattered from the MLC (21). A possible way to improve dose calculation accuracy is to use an algorithm such as Monte Carlo (MC), which can explicitly account for MLC scatter and leakage radiation during the dose calculation simulation. When compared to measurements, MC dose algorithms provide more accurate estimate of a dose delivered to a patient and, consequently, have been used to evaluate dose calculation accuracy (22–29).

Wang et al. (29) investigated the use of MC to evaluate dosimetric effects of inhomogeneities for 5 clinical HN IMRT plans and 5 lung IMRT plans. A measurement-based pencil beam algorithm with an equivalent path length inhomogeneity correction was compared with MC, with the fluence modulation incorporated into the MC using the treatment planning system’s effective MLC transmission matrix. Although the majority of the dose indices in this study agreed well, 2 of the 10 plans showed dose indices with deviations >5%.

Leal et al. (24) studied the use of MC for IMRT verification. Step-and-shoot IMRT plans were calculated using the Plato treatment planning system (Veenendall, Netherlands) and compared with EGS4-based MC calculations and film dosimetry for 3 cases. Major differences were found in heterogeneous situations such as head and neck.

Francescon et al. (22) compared step-and-shoot IMRT dose distributions calculated using the Pinnacle TPS (Philips Medical Systems, Milpitas, CA) collapsed cone convolution algorithm with EGS4-based MC calculations for two plans. In the MC algorithms, particles were simulated through the MLC using the BEAM module, which simplifies the MLC leaf geometry by ignoring MLC interleaf transmission. Isocenter point dose differences of 2.1% for a prostate plan and 2.9% for a head-and-neck plan were observed, and dose–volume histogram (DVH) data comparisons showed deviations of up to 6% for doses below 85% of the prescribed dose and higher deviations for doses above 85% of the prescription dose.

A study by Ma et al. (26) compared IMRT plans from Corvus TPS (Corvus, Nomos Corp., Sewickley, PA) with MC for two patient plans. In the MC simulation, an independently developed effective MLC fluence transmission matrix was used to incorporate intensity modulation into the MC dose calculation. Leaf leakage radiation was included as an empiric correction term to the model. Corvus’s finite-size pencil beam algorithm—calculated doses were found to have discrepancies of >5% for target structures and over 20% for critical structures, compared with the MC results.

The goal of this study was to evaluate the accuracy of previously reported SIB-IMRT dose distributions (11, 13) that were originally computed using SC by comparing results with detailed MC dose computations. To ensure accurate characterization of the radiation transmitted through the MLC during IMRT delivery, the intensity modulation was incorporated into the MC dose calculation using an algorithm that tracked particles through the moving MLC leaves (27). As a result, the MC fully simulated the effect of the MLC on treatment delivery and on radiation transport through the patient. The results of this paper will allow quantitative assessment of the dosimetric accuracy of previously reported SIB-IMRT treatments.

**METHODS AND MATERIALS**

Thirty-one SIB-IMRT plans from 24 different patients who participated in our institutional review board–approved locally advanced phase I/II head-and-neck squamous cell carcinoma protocol were used in this study (11). Details of the plans used for treatment are covered in detail elsewhere (11, 13), but are summarized here for completeness. All plans were created with 9 6-MV equally spaced coplanar beams, with the exception of Plan 28 (Patient 21), in which the location of the target structure led to unequal beam spacing and the use of 3 18-MV photon beams. Four patient plans were created for radiation delivery with a Varian 2100C LINAC (Varian Medical Systems, Palo Alto, CA) equipped with an 80-leaf MLC, and the remaining 27 plans were developed for a Varian 21EX LINAC (Varian Medical Systems, Palo Alto, CA) equipped with a 120-leaf Millennium MLC. Because of the extent of the neck irradiated and the limitation of MLC leaf movement per carriage position, some beams were split into two fields, resulting in 18 ports per plan using a feathered dynamic beam-splitting technique (31) with 2 cm of overlapping beams for the same gantry angle. This technique allowed the LINAC to deliver radiation to half of the field; then the collimator jaws and MLC carriage were moved to irradiate the other half of the field (with a feathered overlap region).

**SIB-IMRT**

Although the SIB treatment technique is described in greater detail by Wu and Mohan (32), and its application with this patient cohort is described in detail by Lauve et al. (11), a brief review is provided here. Target structures used in the SIB optimization include the gross tumor volume (GTV), the clinical target volume (CTV) for subclinical disease, and the electively treated nodal volume (ETV). Methods to obtain target volumes are detailed in Lauve et al. (11). In the protocol, plan dose levels were specified to the GTV, CTV, and ETV and not to corresponding planning target volumes (11, 13, 33, 34); however for this study, this is inconsequential, because the structures used for the dose evaluation were identical for the two dose calculation methods. Three dose levels were used in the SIB protocol. For the first dose level, the GTV received a dose of 68.1 Gy, the CTV received a dose of 60 Gy, and the ETV received a dose of 54 Gy. In the second and third dose levels of this protocol, the GTV received doses of 70.8 Gy and 73.8 Gy, respectively. The dose to the CTV and ETV remained the same for all dose levels. Additional planning goals were to limit the parotid dose to below 30 Gy for at least 50% of the volume, to limit the maximum dose to the spinal cord to 45 Gy, and to limit the maximum dose to the brainstem to 55 Gy. These same dose–volume indices were used for dosimetric evaluation.
VCU-IMRT process

The treatment plans were created using our in-house-developed IMRT planning system (VCU-IMRT). The functionality of VCU-IMRT and the design of SIB-IMRT have been described in earlier publications (21, 31, 32, 35). The VCU-IMRT planning process is as follows: Contours are drawn and approved by the physician, and the dose levels for each structure are specified. Treatment beam directions and angles are then configured using the Pinnacle3 TPS (Philips Medical Systems, Milpitas, CA). Fluence distributions for each beam are then optimized using VCU-IMRT, which reads treatment beam and other plan information directly from the TPS and performs the optimization. During optimization, dose computation occurs within the TPS using Pinnacle’s SC (Adaptive Convolution) algorithm, with intensity modulation specified as a transmission compensator matrix, imported from the VCU-IMRT program. After successful optimization, the optimized transmission compensator matrix is converted to an MLC leaf sequence and a corresponding deliverable effective MLC transmission compensator matrix, which approximately accounts for the effects of head scatter and interleaf and intraleaf radiation leakage on the radiation fluence. This deliverable intensity matrix is loaded into the TPS as a deliverable transmission compensator matrix, and dose is calculated using Pinnacle’s SC dose engine with 4 × 4 × 4 mm³ voxel resolution. Intensity modulation was achieved with the dynamic MLC (sliding window) technique. After planning, MLC leaf sequence files were transferred to the record-and-verify (VARIS; Varian Medical Systems, Palo Alto, CA) system for delivery and subsequently used for film dosimetry.

Monte Carlo

To investigate the accuracy of these SIB-IMRT plans, MC was used to recompute the dose for the 31 plans. The VCU-MC system is based on the EGS4 (36) code, along with user codes BEAM (30) and DOSXYZ (37). One unique ability of the MC system used in this study is that it transports particles through the moving MLC for dose calculation (27). The MC simulation procedure used in this study is as follows: Beam configuration data such as jaw positions, gantry angles, and patient densities, along with any CT density overrides, were directly read from the TPS. In the simulation, particles in each beam were transported from a previously commissioned phase-space (38, 39), which specifies particle positions, directions, and energies on a plane immediately below the ion chamber, through the jaws using BEAM, through the dynamic MLC using our in-house MC code (27), and through the patient using DOSXYZ, where energy deposited was scored. The dose computation grid used for the DOSXYZ MC calculations included the entire CT data set and was 4 mm in each dimension. For each beam, dose was computed to a statistical uncertainty of ~2% at a depth of Dmax. The overall statistical uncertainty within the target volume from the addition of all 9 treatment directions was 1%; thus, statistical uncertainty had a negligible impact on DVH analysis (40, 41). The dose distribution from MC was converted from dose to water before being loaded onto the TPS (42). The MC results were computed in terms of dose per monitor unit; therefore, the monitor units used for the patient’s treatment were for the dose evaluation.

Film measurement comparisons

To benchmark the accuracy of each dose computation algorithm in homogeneous media, each treatment beam from 7 plans (1–3, 21, 29–31) from Patients 1 (2 plans), 2, 15, 22, 23, and 24 was calculated in a homogenous flat water phantom using both SC and MC methods. The results later compared to film measurements at 3 cm depth of a 40 × 40 × 20 cm phantom. The Kodak XV2 (Eastman Kodak Co., Rochester, NY) films were irradiated in a geometry identical to that used for the computations. Films were perpendicular beam central axis with a source-to-film distance of 100 cm. For all film measurements and computations, the gantry angle was set to 0°. Film measurements used the same MLC files exported from the VCU-IMRT planning system as were used for treatment and for the MC dose computations. Films were scanned using Vidar XVR 12+ (VIDAR Systems Corporation, Herndon, VA) film scanner. Film analysis was performed using an in-house film scanning program. Superimposed dose distributions were analyzed for gamma index, dose difference, and distance to agreement based on a published report by Low et al. (43). The fraction of points passing with gamma <1 was calculated using a 2% of the maximum dose and 2 mm distance to agreement criteria.

Plan evaluation

The plans were initially planned and treated using the VCU-IMRT system based on SC dose engine and verified experimentally with film dosimetry. Although isodose display and DVHs give valuable information about the plan, the following end points were evaluated to quantitatively assess each plan.

For the GTV, we evaluated the minimum dose received by 98% of the volume (D98), the maximum dose received by 2% of the volume (D2), the dose received by 50% of the volume (D50), mean dose (Dmean), Equivalent Uniform Dose (EUD) and the Homogeneity Index, which is the ratio of the maximum and minimum doses over the prescription dose. EUD was computed based on a report published by Niemierko (44). The reference dose (Dref) and surviving fraction are 2 Gy and 0.5, respectively.

For the CTV, we evaluated the D98, D50, and Dmean for the ETW, we evaluated the D98, D50, and Dmean. The intended prescription volume coverage levels for the GTV, CTV, and ETW were D98, D50, and D98, respectively. In some of the plans, all elements of the dose prescription were not fully met, because of the inability of the optimization system to achieve all of the criteria. For the parotid glands, we evaluated D98 for right and left glands both separately and together. For spinal cord and brainstem, D2 was evaluated for the maximum dose. The protocol objective was to keep the cord and brainstem D2 doses <45 Gy and <55 Gy, respectively (13). Furthermore, quantities such as Dmax are, by definition, those that differ the most from the mean. In MC methods, it is likely that such an observed value would be the result of statistical uncertainties, whereas a quantity to a prespecified volume such as D2 is less prone to such statistical fluctuations. For the patients in this cohort, the D2 volumes ranged from 0.18 to 0.84 cm³, with an average D2 volume of 0.4 cm³. The dose to 0.01 cm³ (D0.01cm³) was also evaluated.

To effectively compare dose differences between the SC and MC dose calculation methods, comparisons were made relative to the dose computed using the SC algorithm at the local point of interest:

$$\text{Relative percent difference} = \frac{D_{MC} - D_{SC}}{D_{SC}} \times 100$$

And with respect to the treatment prescription:

$$\text{Rx percent difference} = \frac{D_{MC} - D_{Pos}}{D_{Pos}} \times 100$$
These equations yield a positive value if the MC-calculated dose is higher than the SC-computed dose and a negative value if the MC-calculated dose is lower than the SC-computed dose. Each of the dose–volume indices was obtained from the differential DVHs exported from the TPS. Dose differences were tabulated to determine indices and plans that differed by \(3\%\), by \(5\%\), and by \(5\%\).

RESULTS

Film measurements from 7 patient plans were analyzed and are summarized in Fig. 1. The numbers given in the figure are the percentage of points with gamma \(\leq 1\) with a 2\% max dose and 2 mm distance-to-agreement distance criteria. The arithmetic average of all 7 plans with percentage of points passing gamma test is 76.6\% with a range of 65.9--95.8\% for SC and 95.4\% with a range of 91.9--97.2\% for the MC. This result demonstrates that our MC implementation better predicts in-phantom doses and can be used as a benchmark standard. The cause of the improved MC agreement in phantom is the superior intensity-modulated fluence prediction by the MC algorithm, which directly transports particles through the moving MLC.

Patient dosimetric results are presented first for a single sample plan and then summarized for all 31 plans. Table 1 compares the dose–volume-based indices predicted using both the SC and MC dose calculation algorithms for a sample plan (Plan 31). For this plan, the physician-approved prescription was to deliver 70.8 Gy to 92.6\% of the GTV, 60 Gy to 95\% of the CTV, and 54 Gy to 88\% of the ETV. When the plan was recomputed with MC, the predicted doses were higher for each index. The maximum deviation for target structures was observed for the CTV \(D_{50}\) dose (3.4\%, 2.3 Gy). The GTV homogeneity index increased from 5.6\% for SC to 6.5\% for MC. The maximum deviation was found for the right parotid \(D_{50}\) dose, for which the SC dose was 1.9 Gy or 10.4\% lower than that predicted by MC (17.5 Gy vs. 19.4 Gy). However, this amounts to only 2.7\% of the prescription dose and, even with the higher dose, the parotid dose remains within the desired plan criteria. Similar results are found for the cord and the brainstem. Although

<table>
<thead>
<tr>
<th>Index</th>
<th>SC (Gy)</th>
<th>MC (Gy)</th>
<th>Local percent difference (%)</th>
<th>Rx percent difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross tumor volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(D_{98})</td>
<td>70.1</td>
<td>72.3</td>
<td>3.1</td>
<td>3.0</td>
</tr>
<tr>
<td>(D_{\text{Mean}})</td>
<td>72.3</td>
<td>74.6</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>(D_{50})</td>
<td>72.4</td>
<td>74.6</td>
<td>2.9</td>
<td>3.0</td>
</tr>
<tr>
<td>(D_{\text{TVD}})</td>
<td>72.0</td>
<td>74.2</td>
<td>3.0</td>
<td>3.1</td>
</tr>
<tr>
<td>HI</td>
<td>5.6%</td>
<td>6.5%</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Clinical target volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(D_{95})</td>
<td>60.6</td>
<td>62.4</td>
<td>3.0</td>
<td>2.5</td>
</tr>
<tr>
<td>(D_{\text{Mean}})</td>
<td>67.8</td>
<td>70.0</td>
<td>3.3</td>
<td>3.1</td>
</tr>
<tr>
<td>(D_{50})</td>
<td>68.5</td>
<td>70.8</td>
<td>3.4</td>
<td>3.3</td>
</tr>
<tr>
<td>Electively treated nodal volume</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(D_{90})</td>
<td>53.7</td>
<td>54.4</td>
<td>1.1</td>
<td>0.9</td>
</tr>
<tr>
<td>(D_{\text{Mean}})</td>
<td>57.2</td>
<td>58.3</td>
<td>1.9</td>
<td>1.5</td>
</tr>
<tr>
<td>(D_{50})</td>
<td>57.2</td>
<td>58.4</td>
<td>2.1</td>
<td>1.7</td>
</tr>
<tr>
<td>Parotids</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(D_{\text{Mean}})</td>
<td>30.4</td>
<td>32.8</td>
<td>7.6</td>
<td>3.3</td>
</tr>
<tr>
<td>(D_{50})</td>
<td>27.4</td>
<td>29.7</td>
<td>8.5</td>
<td>3.3</td>
</tr>
<tr>
<td>Right parotid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(D_{\text{Mean}})</td>
<td>22.5</td>
<td>24.5</td>
<td>8.9</td>
<td>2.8</td>
</tr>
<tr>
<td>(D_{50})</td>
<td>17.5</td>
<td>19.4</td>
<td>10.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Left parotid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(D_{\text{Mean}})</td>
<td>37.7</td>
<td>40.3</td>
<td>7.0</td>
<td>3.7</td>
</tr>
<tr>
<td>(D_{50})</td>
<td>35.1</td>
<td>37.7</td>
<td>7.2</td>
<td>3.6</td>
</tr>
<tr>
<td>Cord</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(D_{2})</td>
<td>37.4</td>
<td>38.5</td>
<td>2.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Brainstem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(D_{2})</td>
<td>48.3</td>
<td>49.1</td>
<td>1.6</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Abbreviations: SC = superposition/convolution; MC = Monte Carlo.
MC predicted a higher dose, the doses remained within the tolerance levels. Figure 2 shows the sagittal plane isodose distribution for this plan for SC and MC, as well as the absolute dose difference of the two calculation methods with a range of $-4\text{ Gy}$ to $+5\text{ Gy}$. Greater deviations were found in areas where heterogeneity structures are present. Figure 3 shows the DVH data for this plan with the SC dose calculation shown in solid lines and the MC dose calculation in dashed lines. The higher MC dose in all structures for this plan is obvious in this display.

Table 2 shows the average local dose difference, standard deviation, and ranges for the dose indices for all 31 plans. On average, all indices show that MC predicts a higher dose than SC, indicating an $1.5\%$ systematic error in the original SC dose computations. Although the average local dose difference for the GTV $D_{98}$ is only $0.8\%$, deviations as high as $-6.7\%$ (Plan 1) were observed. The maximum deviation for the target structures was seen in the $D_2$ of the GTV ($2.5\%$ with a range of $-4.1\%$ to $+6.1\%$). For the spinal cord and brainstem $D_2$ values, the average local dose difference was $1.9\%$; however, differences of $-5.0\%$ to $6.5\%$ were seen in some plans. Evaluation of the cord $D_{0.01\text{cc}}$ dose revealed an average local difference of $2.3\%$ with a range of $-6.6\%$ to $+8.3\%$, a range similar to that of the $D_2$ value. The main concern with these structures is their respective tolerance levels. For the spinal cord, the maximum amount by which the $D_2$ tolerance level was exceeded was $85\text{ cGy}$ (to $4585\text{ cGy}$). For brainstem, with deviations, all 31 MC computed plans indicated doses below $55\text{ Gy}$.

Note in Table 2 that use of the local dose difference averaged over the plans studied alone can be deceiving; high deviations on both sides of the norm lead to small average deviations. This emphasizes the importance of reporting standard deviations and ranges.

Figure 4 shows local dose difference for the $D_{98}$ of the GTV, the $D_{95}$ for CTV, and $D_{90}$ for ETV. The prescription indices indicate that the deviations were greater for the CTV than the GTV and the ETV for 18 of the 31 plans. For the majority of the plans, MC predicted higher doses than those that had been initially predicted with SC. Figure 5 compares the homogeneity index. Again, for the majority of the plans (28 of 31), MC showed the dose distribution to be less homogeneous than the planned SC dose distribution. Plan 20 showed the largest difference (4.0\%) of the homogeneity index between SC and MC. The average homogeneity index difference from all plans is $1.8\%$ with a range of $-0.6\%$ to $+4.0\%$.

Table 3 summarizes the results in terms of the number of plans that satisfied the set of criteria described below:

- Number of plans with all evaluated dose differences $<3\%$ and $<5\%$ of the local and of the Rx dose. Evaluated dose differences included the $D_{98}$ of the GTV, the $D_{95}$ of the CTV, the $D_{90}$ of the ETV, the $D_{\text{mean}}$ of the right, left, and combined parotid glands, the $D_2$ of the spinal cord, and the $D_2$ of the brain stem.
- Number of plans with target structure indices $D_{98}$ of GTV, $D_{95}$ of CTV, and $D_{90}$ of ETV differing by $<3\%$ and $<5\%$ of the local and of the Rx dose.
- Number of plans with a Parotid $D_{\text{mean}}$ $<3\%$, $<5\%$, and $<10\%$ for the right parotid, left parotid, and combined parotid glands.
Number of plans with the MC spinal cord D$_2$ index exceeding 45 Gy due to deviation from the SC plan.

- Number of plans with the MC brainstem D$_2$ index exceeding 55 Gy due to deviation from the SC plan.

Only 1 out of 31 plans had the SC- and MC-computed dose distributions agree within 3% of the local dose for all of the criteria. When the criteria were expanded to 5%, 4 of 31 plans met the criteria. When only the target structures were examined, 7 and 26 plans met the criteria with 3% and 5% tolerance levels, respectively. For most plans, the parotid gland was the structure with a dose deviation that prevented passing the criteria. Hence, the fraction of plans that passed just the parotid test for 3%, 5%, and 10% was also evaluated. For spinal cord and brainstem, the number of plans that exceeded 45 and 55 Gy is shown, because although differences between SC and MC may be large, if the structure remains below its tolerance dose level, the physician will still consider the plan acceptable.

Agreement was better in terms of the prescription dose, with 4 of 31 plans being within 3% of the prescription dose for all criteria and 28 of 31 being within 5%.

### Table 2. Summary of results for all 31 plans, showing average local dose difference, standard deviation and range, along with average Rx dose difference

<table>
<thead>
<tr>
<th>Structure</th>
<th>Index</th>
<th>Average local percent difference (%)</th>
<th>Standard deviation (%)</th>
<th>Range (%)</th>
<th>Average Rx percent difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross tumor volume</td>
<td>D$_{98}$</td>
<td>0.8</td>
<td>2.8</td>
<td>[−6.7, +4.7]</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>D$_{Mean}$</td>
<td>1.5</td>
<td>2.7</td>
<td>[−5.2, +4.8]</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>D$_2$</td>
<td>2.5</td>
<td>2.8</td>
<td>[−4.1, +6.1]</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>D$_{50}$</td>
<td>1.4</td>
<td>2.6</td>
<td>[−5.2, +4.7]</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>EUD</td>
<td>1.3</td>
<td>2.6</td>
<td>[−5.5, +4.6]</td>
<td>1.3</td>
</tr>
<tr>
<td>Clinical target volume</td>
<td>D$_{95}$</td>
<td>2.2</td>
<td>3.0</td>
<td>[−5.6, +6.9]</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>D$_{Mean}$</td>
<td>2.0</td>
<td>2.6</td>
<td>[−5.0, +4.8]</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>D$_{50}$</td>
<td>1.8</td>
<td>2.6</td>
<td>[−5.3, +4.1]</td>
<td>1.7</td>
</tr>
<tr>
<td>Electively treated nodal volume</td>
<td>D$_{90}$</td>
<td>1.3</td>
<td>3.0</td>
<td>[−4.8, +6.4]</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>D$_{Mean}$</td>
<td>1.5</td>
<td>2.7</td>
<td>[−4.5, +4.3]</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>D$_{50}$</td>
<td>1.2</td>
<td>2.7</td>
<td>[−5.3, +3.6]</td>
<td>0.9</td>
</tr>
<tr>
<td>Parotids</td>
<td>D$_{Mean}$</td>
<td>4.1</td>
<td>4.1</td>
<td>[−4.6, +9.0]</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>D$_{50}$</td>
<td>4.4</td>
<td>4.4</td>
<td>[−6.1, +9.5]</td>
<td>1.6</td>
</tr>
<tr>
<td>Right parotid</td>
<td>D$_{Mean}$</td>
<td>3.7</td>
<td>4.3</td>
<td>[−6.9, +8.9]</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>D$_{50}$</td>
<td>3.5</td>
<td>4.7</td>
<td>[−8.0, +10.4]</td>
<td>1.3</td>
</tr>
<tr>
<td>Left parotid</td>
<td>D$_{Mean}$</td>
<td>5.0</td>
<td>4.7</td>
<td>[−5.1, +13.9]</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>D$_{50}$</td>
<td>4.5</td>
<td>5.1</td>
<td>[−9.1, +13.7]</td>
<td>1.6</td>
</tr>
<tr>
<td>Cord</td>
<td>D$_2$</td>
<td>1.9</td>
<td>3.0</td>
<td>[−5.0, +6.5]</td>
<td>1.0</td>
</tr>
<tr>
<td>Brainstem</td>
<td>D$_2$</td>
<td>1.9</td>
<td>2.8</td>
<td>[−4.8, +5.0]</td>
<td>1.1</td>
</tr>
</tbody>
</table>

- Number of plans with the MC spinal cord D$_2$ index exceeding 45 Gy due to deviation from the SC plan.

- Number of plans with the MC brainstem D$_2$ index exceeding 55 Gy due to deviation from the SC plan.

Only 1 out of 31 plans had the SC- and MC-computed dose distributions agree within 3% of the local dose for all of the criteria. When the criteria were expanded to 5%, 4 of 31 plans met the criteria. When only the target structures were examined, 7 and 26 plans met the criteria with 3% and 5% tolerance levels, respectively. For most plans, the parotid gland was the structure with a dose deviation that prevented passing the criteria. Hence, the fraction of plans that passed just the parotid test for 3%, 5%, and 10% was also evaluated. For spinal cord and brainstem, the number of plans that exceeded 45 and 55 Gy is shown, because although differences between SC and MC may be large, if the structure remains below its tolerance dose level, the physician will still consider the plan acceptable.

Agreement was better in terms of the prescription dose, with 4 of 31 plans being within 3% of the prescription dose for all criteria and 28 of 31 being within 5%.

Fig. 4. Local percent differences for all 31 plans for the D$_{98}$ of the gross tumor volume, D$_{Mean}$ of the clinical target volume, and D$_{50}$ of the electively treated nodal volume. Twenty-three of the 31 plans show Monte Carlo–predicted doses to be higher than superposition/convolution doses for all three prescription indices.

Fig. 5. Homogeneity index of the gross tumor volume for superposition/convolution and Monte Carlo dose calculation algorithms for all 31 plans. The majority of the plans (28 of 31) show the Monte Carlo dose distribution to be less homogeneous than the planned superposition/convolution dose distribution.
In Fig. 6, the histograms of the local differences between the MC and SC dose calculations for the indices are tabulated. For the prescription indices (GTV $D_{98}$, CTV $D_{95}$, and ETV $D_{90}$), 88 of 95 of the indices had deviations of $<5\%$. Although the majority of the plans had deviations below 5%, most of the plans showed MC to be higher than SC.

Figure 6b differentiates the plans into those that required split fields and those that did not. For this case, $D_{50}$ is used in the evaluation, because indices such as $D_{95}$, $D_{\text{mean}}$, and $D_2$ are sensitive to hot and colder spots between the two calculation methods. Split-beam plans showed MC to be higher than SC; similarly, all of the non-split-beam plans showed SC to be higher than MC. The higher MC dose for the split beams indicates that MLC leakage and scatter (which is better predicted by an MC algorithm that explicitly includes radiation transport through the MLC) plays a larger role for these beams.

In Fig. 6c, which histograms the $D_{\text{mean}}$ dose for the parotid glands, 28 of the 31 plans had mean dose deviations $<10\%$, but 3 plans were observed with left parotid deviations $>10\%$. For the spinal cord and brainstem $D_2$ (Fig. 6d), two plans were found to have $D_2$ spinal cord local dose differences $>5\%$. Although the histograms for both spinal cord and brainstem show deviations on both sides of the norm, it is apparent that MC predicts higher doses in the majority of the plans.

**DISCUSSION**

This study shows that MC predicts higher doses than the SC for the dose calculation methods used for this patient cohort. Film dosimetry results confirm that the MC results can be considered to be a benchmark standard. The correlation between phantom dosimetry results and patient results is demonstrated in Fig. 7, which shows the average dose difference with respect to film dosimetry (MC $-$ SC) observed for points with doses $>10\%$ of $D_{\text{max}}$ dose for the phantom calculations and the local percent differences of the GTV $D_{50}$ for the same patients. For each case, the phantom and patient data set deviations were found to be on the same side of the norm, indicating that the phantom-observed deviations are predictive of the patient deviations. However, Fig. 7 also shows that the patient data set deviation is higher than phantom calculation for all but one plan, indicating that additional systematic sources of deviation exist.

**Several potential causes of the differences between the SC and MC results exist, but the three most likely causes are the following:**

- Estimation of fluence upon the patient,
- Dose deposition within the heterogeneous patient media,
- Beam modeling differences between the SC and MC algorithms.

During the estimation of fluence upon the patient, for the SC dose calculation, the intensity modulation is incorporated into dose computation via a transmission compensator matrix. This matrix only approximates the effect of MLC transmission and scatter on the fluence incident upon the patient. The MC algorithm, on the other hand, uses the MLC leaf-sequence file as input and transports the photons directly through the moving MLC leaves (27). Thus, the MC algorithm inherently includes effects from leaf scatter, tongue-and-groove, and beam hardening on the radiation fluence incident upon the patient. For the SC algorithm used in this study, these effects were only approximately incorporated into the fluence modulation matrix, even though our planning system used standard methods to convert leaf trajectories into fluence modulation (45–48). Note that a more accurate fluence-to-trajectory estimator than the one used for these patients would improve the SC algorithm results. Thus, these results should not be construed as a general criticism of the SC dose computation method.

In terms of the effect of heterogeneities, although Fig. 1 shows apparent increases in dose differences near such heterogeneities, indicating that the dose deposition within the heterogeneous patient media is a potential source, previous studies have found that SC compares favorably with MC for HN sites for non-IMRT treatments (49). Isolation of
the fractional contribution of patient heterogeneities to the observed dose deviation is the subject of further study.

Beam modeling is an additional potential cause of deviations. Although the MC algorithm (38, 39, 50) has been commissioned to match measurements, and likewise, the SC algorithm has been clinically commissioned, these two algorithms do use different models of the fluence exiting the accelerator. These different models are potential sources of dose discrepancies, and study of this contribution is also the subject of further investigation.

The fact that MC- and SC-computed SIB-IMRT dose distributions differ implies that the SC plans used for treatment contain convergence errors (35, 40) and that better dose distributions might be possible by using a more accurate dose computation algorithm during plan optimization.

For instance, an MC-optimized plan might allow greater normal-tissue sparing while increasing conformality and the possibility of dose escalation.

Even without detailed knowledge of the direct causes of the discrepancies between the SC treatment calculations and the MC verification calculations, the potential clinical consequences of the observed dosimetric differences can be surmised. For example, some deviations observed in patient doses were greater than the differences between the protocol dose levels (2.3 and 3.0 Gy). For the example patient shown in Table 1, MC doses were higher than SC doses by nearly 2 Gy for all indices. This corresponds with deviations of ~10% of the local dose difference and ~4% of the prescription dose difference. For the prescription dose (GTV \( D_{98} \)), the patient MC-reported dose of 72.3 Gy placed this
accurate patient dosimetry. Such deviations merit a re-evaluation of the previously reported clinical outcome data. For other patients, observed deviations clearly placed the patient in different (higher) dose levels of the protocol. The intended dose level for this patient was 70.8 Gy. For patients shifting to higher dose levels, these deviations can have a substantial impact on dose escalation protocols, resulting in patients shifting to higher dose levels. Furthermore, these results will be vital in correlating structure doses to clinical outcomes, especially for indices in which significant differences were observed, such as the parotid. Overall, these results show that MC clearly can play a significant role in SIB-IMRT. As treatment delivery techniques increase in complexity, the need for an accurate dose calculation algorithm may become a necessity in the everyday radiotherapy clinic.

CONCLUSIONS
Monte Carlo was used to recompute dose distributions for 31 SIB-IMRT plans based on 24 different patients. On phantom, measurements demonstrated that MC dose calculations agree better with film than the SC algorithm as implemented using our in-house IMRT system, justifying its use as reference standard for patient-based dose computations. In the patient geometry, the SC algorithm results from the VCU-IMRT system agreed, on average, with the MC-based algorithm. Over all plans, the arithmetic average of the local percent differences was <2.5% for target structures, <5.0% for the parotid glands, and <1.9% for the spinal cord and brainstem. However, the average results were found to be deceiving: Only 1 out of the 31 plans had dose agreement with respect to the local dose within 3% for all structures, and 4 out of 31 had agreement within 5%. When comparing with respect to the prescription dose, only 4 of the 31 plans agreed within 3%, but most (28 of the 31 plans) agreed within 5% for all criteria. Most plans differed as a result of dose deviations in the parotid glands, which are near the buildup and high-dose gradient regions. The MC dose results were less homogeneous in distribution than those predicted by the SC implementation. Some target structures in some plans had a >5% difference between the SC and MC algorithms. Although the precise causes of these dose deviations remain under investigation, it is important to note that these deviations can have a substantial impact on dose escalation protocols, resulting in patients shifting to higher dose levels. Furthermore, these results will be vital in correlating structure doses to clinical outcomes, especially for indices in which significant differences were observed, such as the parotid. Overall, these results show that MC clearly can play a significant role in SIB-IMRT. As treatment delivery techniques increase in complexity, the need for an accurate dose calculation algorithm may become a necessity in the everyday radiotherapy clinic.

REFERENCES
12. Butler EB, Teh BS, Grant WH, 3rd, et al. Smart (simultaneous modulated accelerated radiation therapy) boost: A new accelerated fractionation schedule for the treatment of head and


