EVALUATION OF IMAGE-GUIDANCE PROTOCOLS IN THE TREATMENT OF HEAD AND NECK CANCERS

OMAR A. ZEIDAN, PH.D., KATJA M. LANGEN, PH.D., SANFORD L. MEeks, PH.D., RAFAEL R. MANON, M.D., THOMAS H. WAGNER, PH.D., TWYLA R. WILLOUGHBY, M.S., D. WAYNE JENKINS, M.D., AND PATRICK A. KUPELIAN, M.D.

Department of Radiation Oncology, M. D. Anderson Cancer Center Orlando, Orlando, FL

Purpose: The aim of this study was to assess the residual setup error of different image-guidance (IG) protocols in the alignment of patients with head and neck cancer. The protocols differ in the percentage of treatment fractions that are associated with image guidance. Using data from patients who were treated with daily IG, the residual setup errors for several different protocols are retrospectively calculated.

Methods and Materials: Alignment data from 24 patients (802 fractions) treated with daily IG on a helical tomotherapy unit were analyzed. The difference between the daily setup correction and the setup correction that would have been made according to a specific protocol was used to calculate the residual setup errors for each protocol.

Results: The different protocols are generally effective in reducing systematic setup errors. Random setup errors are generally not reduced for fractions that are not image guided. As a consequence, if every other treatment is image guided, still about 11% of all treatments (IG and not IG) are subject to three-dimensional setup errors of at least 5 mm. This frequency increases to about 29% if setup errors >3 mm are scored. For various protocols that require 15% to 31% of the treatments to be image guided, from 50% to 60% and from 26% to 31% of all fractions are subject to setup errors >3 mm and >5 mm, respectively.

Conclusion: Residual setup errors reduce with increasing frequency of IG during the course of external-beam radiotherapy for head-and-neck cancer patients. The inability to reduce random setup errors for fractions that are not image guided results in notable residual setup errors. © 2007 Elsevier Inc.

INTRODUCTION

The increased use of conformal irradiation techniques has been accompanied by the development of image-guidance (IG) techniques to improve patient positioning (1–6). All IG techniques are designed to integrate into the treatment process, and the ease of operation of some of the developed techniques allows their daily use. However, each technique is still associated with cost in terms of machine time, patient imaging dose, or both. This cost must be evaluated in the context of improved target positioning.

It is clear that the daily use of an accurate and precise IG technique results in the best possible patient positioning over the course of treatment. Less frequent use of IG will reduce the associated costs at the expense of reduced setup accuracy. The gain in setup accuracy that is expected with the daily use of IG will depend on the nature and magnitude of setup and organ-motion uncertainties. Targets that are subject to interfraction organ motion, such as the prostate, will benefit most from the daily use of IG. However, the gain is less obvious for targets that are less mobile.

Targets in the head-and-neck region are, in general, assumed to be fairly rigidly attached to bony anatomy; and the main source of uncertainty in positioning these patients is setup error. Typically, the setup uncertainty is minimized with head or head-and-shoulder masks that help to reposition and to immobilize the patient. The improvement in setup accuracy that can be achieved for immobilized head-and-neck patients with daily IG is not clear. It is the aim of this study to assess the residual setup errors in immobilized head-and-neck patients for imaging protocols that use IG less often than daily. Using the daily alignment data from 24 head-and-neck cancer patients, eight protocols were retrospectively replayed. The protocols range from using no IG

Acknowledgments—We thank Women Playing for T.I.M.E for their support of our research efforts.

Conflict of interest: none.

Received April 19, 2006, and in revised form Sept 20, 2006. Accepted for publication Sept 26, 2006.

Image guidance, Head-and-neck cancers, Megavoltage CT, Alignment protocols, TomoTherapy.
to using IG with every other fraction. The residual setup errors will quantify the improvement in setup accuracy for this class of patients that can be achieved with daily IG.

**METHODS AND MATERIALS**

At the M. D. Anderson Cancer Center Orlando, a commercial helical tomotherapy unit (Hi*Art II; TomoTherapy Inc., Madison, WI) (7) has been used to treat patients with head and neck cancer. Since October 2003, approximately 350 patients have been treated with this modality. In our current clinical practice, all patients treated are imaged using megavoltage CT (MVCT) scans before each fraction. These MVCT scans are used to correct patient position before each treatment. Daily alignment data were available for 24 head-and-neck cancer patients treated with helical tomotherapy. The number of fractions per patient ranged from 20 to 35. A total of 802 image-guided treatments were available for analysis, with an average of 33 treatments per patient. In all patients, immobilization was achieved by use of a head mask (MEDTEC, Orange City, IA), which was firmly attached to the treatment couch in the same position for all patients.

**Planning and daily alignments**

For treatment-planning purposes, a kVCT scan (3-mm slice thickness) was obtained of each patient. In preparation for treatment, each patient was initially aligned using room lasers and permanent marks on the patient mask. A megavoltage (MVCT) scan was subsequently obtained (3). The MVCT scans are obtained with a normal or coarse pitch setting that corresponds to nominal slice thicknesses of 4 and 6 mm, respectively. After the MVCT is acquired, it is registered with the kVCT image to determine corrections to the patient position. Typically, a bony anatomy–based automatic fusion algorithm is used. This algorithm is based on an extracted feature fusion algorithm (8). After this initial registration, the radiation therapists visually inspect the images and have the option of manually modifying the registration. On days when the bony anatomy is deformed relative to the kVCT, optimal alignment throughout the entire region of interest is not possible. Priority is given to the registration of bony anatomy in the high-dose area, i.e., the bony anatomy near the target areas.

**Alignment protocols**

To determine the accuracy and precision of the patient alignment if IG is used less often than daily, several protocols were retrospectively replayed using the daily alignment data collected from the 24 patients.

**Protocol A: no imaging**

In this protocol, each patient’s position is corrected only for a machine-specific systematic positioning error. On our TomoTherapy unit, a systematic drop in the couch height exists between the patient setup location at the virtual isocenter (outside the bore) and the treatment isocenter (inside the bore). The magnitude of this systematic drop was estimated from the first 100 patients treated using the machine. The systematic couch shift is 8 mm in the anterior–posterior (A/P) direction and 2 mm in the superior–inferior (S/I) direction inferiorly. The systematic shift in S/I is related to the A/P drop, as the drop is not unidirectional. Using this protocol, the patient would never be imaged, but after alignment with the room lasers, the systematic correction in the A/P and S/I direction would be performed before treatment. The residual positioning errors that would have occurred with this protocol can be determined by subtracting the systematic corrections from the daily image-guided positioning data. Since no MVCT images are acquired, this protocol has the least workload.

**Protocols B1, B3, B5, and B7: mean shift of initial one, three, five, or seven fractions**

The patient undergoes imaging for the first one, three, five, or seven fractions. For these initial fractions, the patient position is corrected according to the MVCT-kVCT registration. After one, three, five or seven fractions, a mean shift is calculated and applied to each subsequent fraction. No further images are obtained. Four different protocols were investigated using this strategy. These protocols are similar to the no-action–level protocol proposed by de Boer and Heijmen, except that our protocols call for the patient position always to be corrected on days when an MVCT image is obtained (9).

**Protocol C: weekly imaging with 3-mm threshold**

This protocol is designed to mimic the common clinical practice of weekly portal imaging. The patient undergoes imaging before the first fraction. The shifts are applied to the first and all subsequent treatments. Weekly images are acquired. If any shift according to the MVCT-kVCT registration is ≥3 mm, the patient position is adjusted for this shift that day and for all subsequent fractions. If the shift is <3 mm, the daily shifts are not changed.

**Protocol D: first five fractions, then subsequent weekly imaging with patient-specific threshold**

The patient undergoes imaging on Days 1 to 5 and is aligned according to the MVCT-kVCT registration. A mean shift is calculated from these data and is applied on all subsequent treatment days. From the corrections of the first five fractions, the standard deviation of the shifts is calculated for each direction. Imaging is subsequently performed on fraction numbers 10, 15, 20, and so on. Because the patient shifts are available on these days, the patient is properly aligned on Days 10, 15, 20, and so on. The initial shifts calculated from the first five fractions are not changed unless any patient correction is more than two times the standard deviation of the shifts from the first 5 days. If the patient shift is >2 standard deviations, the subsequent fractions are corrected by this new additional systematic shift for all subsequent fractions.

**Protocol E: imaging every other fraction with a running mean**

The patient undergoes imaging every other day. A running average shift is determined, and the patient’s position is corrected by this running average on days when no image is obtained.

**Frequency of IG**

To determine the workload of each protocol, the percentage of treatments that would have been imaged is scored. Using the patient data from the current study, in protocol B with a = 7,168 (7 × 24) of the 802 treatments would have been image guided. This corresponds to a frequency of 21%. Although not each treatment is imaged guided, each fraction is still corrected according to the different IG protocols. For example, in protocol A (no imaging), the imaging frequency is 0%, but a systematic correction is still applied at each treatment fraction. In protocol B, the mean shift of the initial fractions is calculated, and subsequently all patient positions are corrected by this average correction. In prac-
tic, these shifts would be implemented after the laser-based patient alignment. Therefore, whereas not each fraction is image guided, each fraction is nevertheless corrected. This distinction is important, because it is obvious that with an increased imaging frequency, the residual setup error averaged over all fractions decreases. However, the value of any protocol can also be evaluated by separately examining the treatments that are not image guided. A protocol that is most effective in reducing the residual positioning errors for treatments that are not image guided may be preferable.

Data analysis

The difference between the daily alignment and the alignment that would have been made had a particular image-guidance protocol been followed is calculated for each treatment day and IG protocol. These data represent the residual setup errors that would have occurred if the protocols had been used during the treatment. To score the residual alignment errors in each protocol, population-based parameters are calculated based on the methodology introduced by van Herk (10). The residual setup error for each fraction is calculated from the difference between the actual daily alignment and the alignment according to a given protocol. For each patient, the systematic setup error, M, and the standard deviation (sd) can be calculated for each protocol averaged over all fractions. A population-based mean systematic setup error, m, is calculated by averaging the systematic setup errors, M_{1−24}, from all patients. The standard deviation of the systematic error, \( \sigma \), is equal to the standard deviation of the patient-specific systematic errors, M_{1−24}. The average random error over the whole patient population, \( \sigma \), is determined by calculating the root-mean-square of the random setup errors (sd) observed for each patient. In addition, the frequency of the residual positional errors can be determined for each protocol. Population-based statistics and the frequency of residual setup errors were separately calculated for fractions that were not associated with IG to test the predictive quality of each protocol.

Accuracy and precision of image-guided alignments

For treatments that are accompanied by IG, it cannot necessarily be assumed that the residual positioning errors are zero after MVCT-kVCT image registration. To determine the accuracy and precision of the patient setup after MVCT-kVCT image registration, the alignment process was tested using a rigid wax-filled head phantom. A kVCT scan of this phantom was obtained. To test the precision of the MVCT-kVCT alignment, the phantom was aligned at the virtual isocenter using a laser and fiducial markers. The phantom was then moved by known amounts using table shifts and rescanned. The second MVCT scan was again registered with the kVCT and the table position was corrected accordingly. The resulting table positions were compared with the reference table position to determine the reproducibility of the MVCT to kVCT registration process. This was repeated 10 times to determine the precision of the alignment. The standard deviation of the final table position was 0.2, 0.7, and 0.3 mm in the anterior–posterior, superior–inferior, and lateral direction, respectively. The values can be used to estimate the average random error, \( \sigma \).

RESULTS

A summary of the alignment data from the daily image-guided sessions is provided in Table 1. These data show the daily setup corrections that were performed for the 24 patients (802 fractions). The systematic couch drop in the anterior–posterior direction and the related systematic couch shift in the superior–inferior direction is manifested in the 8.72 and −2.8 mm respective mean systematic setup corrections, M. The mean random corrections are in the range of 2 to 3 mm. These are in agreement with random setup errors reported in the literature for immobilized head-and-neck patients (11).

Figure 1 shows the mean systematic error, M; the standard deviation of the systematic error, \( \sigma \); and the average random error, \( \sigma \), for the residual setup errors after each protocol is applied retrospectively to the daily setup corrections. The data are presented for all treatments (all TX), either image guided or not. In addition, the data for the fractions that were not image guided (non-IGRT) are shown. The data are plotted against the imaging frequency associated with each protocol. An analysis of Fig. 1 shows that all protocols are effective in reducing the mean systematic error, m, to less than 1.5 mm. The patient-to-patient variation of the systematic error, \( \sigma \), does, in general, reduce with increasing imaging frequency. This component is reduced to \( <1.5 \) mm (averaged over all fractions) in all directions for protocols with imaging frequencies of 9% or more. A reduction of this variation to \( \sim0.5 \) mm in all directions is observed in the protocol that uses IG in every other fraction (imaging frequency 50%). If only the treatments that are not image guided are analyzed, the patient-to...

| Table 1: Mean systematic setup correction, M; standard deviation of the systematic correction, \( \Sigma \); and average random correction, \( \sigma \), for the daily setup corrections after daily image guidance was performed on 24 patients (802 fractions, average 33 per patient) |
|---------------------------------|--------|--------|--------|
| Direction                      | M (mm) | \( \Sigma \) (mm) | \( \sigma \) (mm) |
| Anterior–posterior             | 8.7    | 2.2    | 2.3    |
| Superior–inferior              | −2.8   | 2.6    | 2.2    |
| Lateral                        | 0.0    | 1.6    | 2.3    |
patient variation of the systematic error is larger but generally decreases with increasing imaging frequency. This means that more frequent imaging reduces the systematic errors in the nonimage-guided treatments.

In contrast to systematic errors, random errors are more problematic. Random errors can be secondary to a multitude of technical, patient, or treatment factors. The mean random error in patient setup (averaged over all treatments) decreases with increasing imaging frequency, but no reduction in the random error is seen if the nonimage-guided fractions are examined. The frequencies of errors that are larger than some tolerance are of interest clinically and allow an intuitive judgment of the patient alignment accuracy. The frequency distribution of residual errors is plotted for all treatment fractions and protocols in Fig. 2. This analysis is repeated in Fig. 3 for the fractions that were not image guided. If all treatments are evaluated, most imaging protocols with frequencies of 15–31% result in misalignment >5 mm in the anterior–superior direction in about 10% of the fractions. In the superior–inferior and lateral direction the frequency ranges from about 5–12% and 5–18%, respectively. However, with protocol E (imaging every other fraction with a running mean), which has an imaging frequency of 50%, the rate of such errors is reduced to 2–3% in each direction. If a 3-mm margin is used, the frequency of partial misalignments is still about 9% in all directions if protocol E (imaging every other fraction with a running mean) is used.

If only treatments are evaluated that are not accompanied by IG (Fig. 3), the frequency of residual setup errors varies less with different IG protocols. The magnitude of the three-dimensional displacement can be calculated from the residual directional setup errors. The frequency of this vector is plotted in Fig. 4 for all treatments and the nonimage-guided treatments. Table 2 lists the percentage of all treatments in which the three-dimensional residual setup errors exceed 3, 5, and 10 mm. For protocols that have an imaging frequency between 15% and 31%, the percentage of treat-
ments that are subject to residual errors $\geq 3$ mm and $5$ mm ranges from $50\%$ to $60\%$ and $26\%$ to $31\%$, respectively. If every other treatment is image guided and the nonimage-guided treatment is corrected by a running average, still about one-third of all treatments is associated with a setup error of at least 3 mm. Using this protocol, about 10% of all treatments are limited by setup errors of at least 5 mm. The percentage of IG for each protocol is shown to illustrate the tradeoff between residual error frequency and the imaging workload per protocol.

**DISCUSSION**

The availability of daily alignment data allows the evaluation of different image-guidance protocols. The gain in alignment accuracy and precision that is achieved if daily IG is used can thus be assessed. Furthermore, systematic and random contributions to the residual setup uncertainty can be calculated. The above findings, namely, that imaging protocols are only effective in reducing the systematic component of the setup errors, are in agreement with expectations. This systematic component can be better estimated with more frequent imaging. The random error component is only reduced if averaged over all fractions. No reduction in the random error is seen for the fractions that are not image guided. Residual setup errors in each IG strategy are mostly the result of the random position fluctuations.

The frequency of residual errors allows a judgment of the clinical impact of each imaging protocol. The frequency of setup errors that can be tolerated is a clinical decision that may depend on the specific treatment site and patient-specific conditions. The ability of each protocol to reduce the frequency of setup errors when IG is not performed can be deduced from Fig. 3. In each direction, Protocol D (IG in the first five fractions and weekly thereafter) fairs among the worst in terms of reducing setup errors. Not surprisingly, the two protocols with the lowest imaging frequency are also the least effective in reducing setup errors when IG is not used. Conversely, the protocol that calls for imaging at every other fraction (i.e., highest imaging frequency evaluated) is best at reducing setup errors on days when IG is not performed. If the magnitude of setup errors is evaluated as a three-dimensional vector, the residual setup errors are obviously larger than if the three dimensions are evaluated separately. From Fig. 4a and Table 2 it can be seen that with an imaging frequency of 50%, nearly 11% of all treatments are subject to a three-dimensional setup error of $>5$ mm. For the same protocol about one-third of all treatments is limited by setup errors $>3$ mm.

It is clear that an improvement in setup accuracy and precision can be achieved if each fraction is accompanied
by IG. The phantom tests show submillimeter accuracy and precision of the MVCT to kVCT registration process. However, it must be kept in mind that the skull phantom is a rigid body and presents a best-case scenario for MVCT to kVCT image registration. In reality, patient anatomies may deform during the course of treatment, and bony anatomy registration in the head-and-neck region is frequently compromised by changes in the patient’s anatomy. A margin to account for deformation and other uncertainties in the delivery process is necessary and must be added irrespectively of setup.

Fig. 3. The frequency of residual setup errors is shown for all protocols. Only treatments that are not associated with image guidance are included in this analysis. (a–c) Results for the three directions. IGRT = image-guided radiation therapy; TX = treatment.

Fig. 4. The frequency of residual three-dimensional setup errors is shown for all protocols. (a and b) Results for all treatments and the nonimage-guidance (IG) treatments. IGRT = image-guided radiation therapy; TX = treatment.
dose is distributed throughout the imaged patient anatomy. Contrary to portal imaging, MVCT without using a normal or coarse pitch setting, the doses were 0.7 and 0.4 cGy, respectively (13). In our clinical practice the acquisition of the MVCT registration cannot account for this uncertainty, and image registration, and adjustment of patient alignment in the head-and-neck region. Furthermore, patients may move during the actual treatment. The MVCT to kVCT registration cannot account for this uncertainty, and its effect must be minimized with the use of appropriate PTV margins.

The daily use of MVCT guidance is associated with an increased patient imaging dose as well as the machine time per patient. In our clinical practice the acquisition of the MVCT, image registration, and adjustment of patient alignment adds about 5 to 10 min to the in-room time. In comparison, the typical beam-on time for head-and-neck treatments is <10 min. To estimate the dose from MVCT imaging dose measurements were performed in the center of a 20-cm cylindrical acrylic phantom. For images acquired with a normal or coarse pitch setting, the doses were 0.7 and 0.4 cGy, respectively (13). Contrary to portal imaging, MVCT dose is distributed throughout the imaged patient anatomy. This means that anatomy that is removed from traditional beam ports will receive imaging doses that may not have been given if traditional portal images had been used. However, this is not likely to be the case in head-and-neck cancer patients in whom the entire anatomy is included in traditional double-exposure portal images.

The benefit of daily IG for head-and-neck cancer patients can be assessed from our results. Reducing the image-guidance frequency from daily to 50% results in 11% of all treatments to be subject to a misalignment error of at least 5 mm. Approximately one-third of all treatments are subject to a ≥5-mm misalignment if protocols are used that use IG on a weekly basis. The effect of these misalignments on the target coverage can be mitigated if appropriate PTV margins are used. For patients with target volumes in close proximity to critical sensitive structures, PTV margins may have to be small. In these cases daily IG should be used.

**CONCLUSION**

Systematic errors throughout the course of treatment can be reduced with protocols that involve infrequent imaging. However, random setup errors are not eliminated with any of the imaging protocols investigated in the current study. Protocols that use IG for 0–50% of all treatments have three-dimensional residual setup errors of ≥5 mm in 47–11% of all treatment. The acceptability of the reduced setup accuracy and precision that is caused by infrequent use of IG depends on treatment margins and the proximity of sensitive structures. If target structures are in close proximity to critical sensitive structures, IG should be used daily.

**REFERENCES**

4. Jaffray DA, Drake DG, Moreau M, et al. A radiographic and tomographic imaging system integrated into a medical linear...


