IS THERE A SELECTION BIAS IN RADIOTHERAPY DOSE-ESCALATION PROTOCOLS?

ELISABETH WEISS, M.D.,*† VISWANATHAN RAMAKRISHNAN, PH.D.,‡ AND PAUL J. KEALL, PH.D.*§

*Department of Radiation Oncology, Virginia Commonwealth University, Richmond, VA; †Department of Radiation Oncology, University of Göttingen, Göttingen, Germany; ‡Department of Biostatistics, Virginia Commonwealth University, Richmond, VA; and §Department of Radiation Oncology, Stanford University, Stanford, CA

Background: To investigate the existence of a selection bias using a virtual radiotherapy dose-escalation trial. In dose-escalation trials, normal tissue constraints generally remain constant while the tumor dose is increased. Since tumor dose and normal tissue constraints are competing demands, a point will be reached at which the tumor dose cannot be increased without exceeding normal tissue constraints.

Methods and Materials: In 9 patients with non–small-cell lung cancer, the tumor dose was escalated from 66 Gy to 78 Gy in 4-Gy dose levels using intensity-modulated radiotherapy planning, while the limiting normal tissue dose contraints remained constant. Dosimetric, radiobiologic, and other planning parameters were compared at the 66-Gy dose level for patients eligible for all dose levels and for those eligible only for lower dose levels.

Results: Seven of 9 patients were eligible for all dose levels (Group E). Two of 9 patients (“ineligible” or Group I) qualified only for lower total doses (95% confidence interval, 0.075–0.6, significant). In Group E, mean planning target volumes were smaller (132 vs. 404 cm³, nonsignificant), monitor units per fraction were significantly lower (448 vs. 802, p = 0.0008), and the average composite score for plan quality was better than in Group I (0.012 vs. 0.068, nonsignificant). Average tumor-control probabilities were higher (0.33 vs. 0.23, nonsignificant), and normal tissue-complication probabilities were lower for Group E than for Group I.

Conclusions: Patients eligible for higher dose levels had significantly superior estimated outcome parameters. A method to eliminate this source of error in the interpretation of dose-escalation trials is suggested.

Dose escalation, Lung cancer, Normal tissue toxicity, Intensity-modulated radiotherapy, Selection bias.

INTRODUCTION

Increasing tumor doses in radiotherapy (RT) are associated in general with better tumor control. Normal tissue toxicity is considered the major factor limiting higher doses to the tumor. To increase conventionally accepted doses under controlled circumstances, dose-escalation trials are performed. More and more data on the dose- and volume-dependence of normal tissue toxicity are becoming available, particularly in the thorax (1–6), and are incorporated as limiting factors in dose-escalation trials. Recent dose-escalation trials in patients with non–small-cell lung cancer (NSCLC), using conformal, three-dimensional treatment technology, showed that giving higher radiation doses can be achieved safely (2, 7–12). While phase I dose-escalation trials do not provide any insights on the impact of higher doses on treatment outcome, their results are relevant as a basis for phase II and phase III trials.

It was shown that the risk for pneumonitis (as the major toxicity in thoracic radiotherapy) depends on the lung volume exceeding a threshold dose of 20 Gy (3). Current studies therefore stratify patients according to irradiated lung volume and the estimated pneumonitis risk, allowing higher dose levels only in patients with at low risk, i.e., usually in tumors with a smaller volume (2, 7, 8, 12). Stratification reduces the issue of tumor size, tumor location, and tumor stage as confounding factors in the interpretation of dose-escalation trials. However, wide dose ranges in each stratification group might still introduce a selection bias by virtue of the fact that less favorable tumors will not be eligible for higher dose levels. Although normal tissue dose limitations have repeatedly been identified as a factor limiting dose escalations in individual patients, to the best of our knowledge, no systematic analysis involving clinical data was previously performed.

Acknowledgments—The authors acknowledge Drs. Lei Dong and Radhe Mohan from the University of Texas M.D. Anderson Cancer Center for supplying the CT data used for this study. Thanks go to Devon Murphy for improving the clarity of the manuscript. Received Sept 27, 2006, and in revised form Dec 30, 2006. Accepted for publication Feb 6, 2007.

Reprint requests to: Elisabeth Weiss, M.D., Department of Radiation Oncology, Virginia Commonwealth University, 401 College St., PO Box 980058, Richmond, VA 23298. Tel: (804) 828-9463; Fax: (804) 828-6042; E-mail: eweiss@mcvh.vcu.edu

This study was partially supported by grant R01 CA 93626 from the National Cancer Institute.

Conflict of interest: none.
To test the hypothesis of the existence of a selection bias, a virtual dose-escalation trial was performed in patients with NSCLC. Dose planning was performed with the same increasing dose levels and constant normal tissue constraints for all patients. The aim was to determine whether all patients were eligible for all dose levels, and if not, to identify significant differences in clinical, dosimetric, and radiobiologic parameters between those patients who met or did not meet the required normal tissue dose limitations at any of the given dose levels.

**METHODS AND MATERIALS**

Based on the computed tomography (CT) data sets of 9 sequentially chosen patients with inoperable NSCLC, a virtual dose-escalation study was performed. Tumor location and tumor extent varied between patients. Similar to other dose-escalation studies (Table 1), a large variety of tumor stages was included in this study. Locoregional tumor stages ranged from T1N0 to T3N3, with both peripheral (2 patients) and central (7 patients) tumor locations. The tumors were situated in the upper or middle lobes in 7 of 9 patients. Enlarged mediastinal lymph nodes were diagnosed in 6 patients.

Table 1. Dose-escalation trials in non–small-cell lung cancer

<table>
<thead>
<tr>
<th>Authors (with reference number)</th>
<th>No. of patients, tumor stage, tumor volume</th>
<th>Chemotherapy</th>
<th>Doses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belderbos et al., 2003 (7)</td>
<td>55 patients, Stage I–IIIB</td>
<td>Escalation from 54.0 to 101.3 Gy</td>
<td></td>
</tr>
<tr>
<td>Bradley et al., 2005 (2)</td>
<td>179 patients, Stage I–III; median GTV: Group I, 31.8 cm³; Group II, 79.5 cm³</td>
<td>1 in 25 patients</td>
<td>Group I, 70.9, 77.4, 83.8, 90.3 Gy; Group II, 70.9, 77.4, 83.8 Gy; Group III, 64.5, 70.9, 77.4 Gy (2.15 Gy)</td>
</tr>
<tr>
<td>Kong et al., 2005 (8)</td>
<td>106 patients, Stage I–III</td>
<td>1 in 19% patients (Stage III)</td>
<td>3D CRT, 63.0–102.9 Gy (2.1 Gy)</td>
</tr>
<tr>
<td>Rosenman et al., 2002 (9)</td>
<td>62 patients, Stage IIIA + B; median GTV: Group I, 64 cm³; Group II, 94 cm³</td>
<td>I + C</td>
<td>PTV&lt;sub&gt;prechemo&lt;/sub&gt; 46 Gy appa, PTV&lt;sub&gt;postchemo&lt;/sub&gt; boost to 60, 66, 70, and 74 Gy</td>
</tr>
<tr>
<td>Rosenzweig 2005 (10)</td>
<td>104 patients, Stage I–IIIB</td>
<td>1 in 16% of patients (only at 81- and 84-Gy dose level)</td>
<td>3D CRT, 70.2 Gy (1.8 Gy), 75.6 Gy (1.8 Gy), 81 Gy (1.8 Gy), 84 Gy (2 Gy); 90 Gy (2 Gy) only for T1/2 N0 tumors, but lower f&lt;sub&gt;dam&lt;/sub&gt;</td>
</tr>
<tr>
<td>RTOG 0117</td>
<td>Stage I–IIIB</td>
<td>C</td>
<td>3D CRT arm I, 75.25 Gy (2.15 Gy); arm II, 74 Gy (2 Gy)</td>
</tr>
<tr>
<td>Socinski et al., 2004 (11)</td>
<td>29 patients, Stage IIIA + B; median GTV: Group I, 28 mL; Group II, 66 mL</td>
<td>I + C</td>
<td>PTV&lt;sub&gt;prechemo&lt;/sub&gt;, 40–50 Gy appa, PTV&lt;sub&gt;postchemo&lt;/sub&gt; boost to 78/82/86/90-Gy oblique fields</td>
</tr>
<tr>
<td>Wu et al., 2003 (12)</td>
<td>50 patients, Stage II–IIIB; mean GTV, 63 cm³ (range, 32–308)</td>
<td>I + A</td>
<td>42 Gy (2 Gy) appa, 3D CRT to 69, 72, 75, 78 Gy (3 Gy)</td>
</tr>
</tbody>
</table>

(Continued)

**Abbreviations:** 3D = three-dimensional; A = adjuvant chemotherapy; appa = anteroposterior/posteroanterior; C = concurrent chemotherapy; CLGB = Cancer and Leukemia Group B; CRT = conformal radiotherapy; CTC = Common Toxicity Criteria; CTV = clinical target volume; f<sub>dam</sub> = parallel model fraction damaged parameter; G = grade; GTV = gross tumor volume; I = induction chemotherapy; LN = lymph nodes; MTD = maximum tolerated dose; NTCP = normal tissue complication probability; PTV = planning target volume; RTOG = Radiation Therapy Oncology Group; SWOG = South Western Oncology Group; V<sub>eff</sub> = effective volume; V<sub>x</sub> = volume of a specified organ receiving a dose of x Gy or more.
The gross tumor volume (GTV) was defined as all macroscopically identifiable tumor, including lymph nodes with a diameter of ≥1 cm in the short axis on CT. To obtain the clinical tumor volume (CTV), the GTV was expanded by an 8-mm margin toward the lung tissue (13) and a 5-mm margin around the affected lymph nodes. The CTV contour was adjusted so as not to cross lobe boundaries, and to exclude the chest wall or organs situated in the mediastinum, unless infiltrated by the tumor. Unaffected lymph node areas were not included in the CTV. For the planning target volume (PTV), an 8-mm margin was added isotropically to the CTV to account for internal motion and setup errors. These margins are similar to those in other dose-escalation studies, with or without respiration management (Table 1).

Intensity-modulated radiotherapy (IMRT) planning was performed using the Pinnacle treatment-planning system (version 6.2) with the collapsed-cone convolution implementation of the superposition algorithm (14). All planning was performed for gated delivery during unforced end-expiration, because expiration is the most stable respiratory position with the longest duration in the breathing cycle (15–17), and gated IMRT for lung cancer is in common use at our institution (18). The prescribed dose was increased in 4-Gy dose levels from 66 Gy as a standard dose for NSCLC treatment to 70 Gy, 74 Gy, and 78 Gy, with a fractionation of 5 × 2 Gy/week. The objective of inverse planning was to deliver the prescribed dose to ≥95% of the PTV. In general, six coplanar, nonopposed 6-MV beams with gantry angles 0°, 20°, 40°, or 60° from the anterior or posterior direction were chosen. Beam angles were selected depending on the tumor location. The most stable respiratory position with the longest duration in the breathing cycle (15–17), and gated IMRT for lung cancer is in common use at our institution (18). The prescribed dose was increased in 4-Gy dose levels from 66 Gy as a standard dose for NSCLC treatment to 70 Gy, 74 Gy, and 78 Gy, with a fractionation of 5 × 2 Gy/week. The objective of inverse planning was to deliver the prescribed dose to ≥95% of the PTV. In general, six coplanar, nonopposed 6-MV beams with gantry angles 0°, 20°, 40°, or 60° from the anterior or posterior direction were chosen. Beam angles were selected depending on the tumor location. The initially chosen beam arrangement for the lowest dose level remained the same for all other dose levels, to avoid subjectivity in plan generation and comparison.

While plan optimization was performed using objectives that were adjusted to the changes in total dose, e.g., allowing a dose range of −10% and +20% of the prescribed dose (19, 20), normal tissue constraints were absolute values and remained identical for all dose levels. The total lung volume receiving ≥20 Gy (V20) was

<table>
<thead>
<tr>
<th>Safety margins</th>
<th>Normal tissue dose limits</th>
<th>MTD definition and results</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTV = GTV + 1–1.5 cm, no elective LN</td>
<td>Five risk groups according to relative mean lung dose and NTCPs: Cord &lt;50 Gy, whole heart &lt;40 Gy, 1/3 heart &lt;66 Gy, esophagus Veff &lt;30% for 80 Gy</td>
<td>Lung toxicity ≥G3 (SWOG), acute esophagitis ≥G3, late esophagitis ≥G2 (RTOG), any ≥G3 toxicities (RTOG) MTD not reached, 7 patients ineligible for planned dose step because of normal tissue dose limits (1 cord, 6 esophagus)</td>
</tr>
<tr>
<td>PTV = GTV + 1 cm or more, no elective LN</td>
<td>Group I, Lung V20 &lt;25%; Group II, Lung V20 25–36%; Group III, Lung V20 &gt;36%</td>
<td>RTOG ≥G3 toxicity in ≥15%; MTD Group I, 83.9 Gy; Group II, 77.4 Gy</td>
</tr>
<tr>
<td>CTV = GTV + 0.5 cm, PTV = CTV + 1–1.5 cm, no elective LN</td>
<td>Escalation according to five effective lung-volume bins and NTCPs</td>
<td>Dose predictive for outcome</td>
</tr>
<tr>
<td>PTV = CTV + ≥1.5 cm, CTVprechemo = GTVprechemo + elective LN, CTV boost = GTVpostchemo + 1 cm</td>
<td>Cord &lt;50 Gy, left ventricle ≤40 Gy, lung V20 ≤35%, bronchial plexus &lt;66 Gy</td>
<td>CLGB G3–4 nonhematological toxicity, RTOG for esophagitis, MTD not reached</td>
</tr>
<tr>
<td>Start with elective LN, later PTV = GTV + 1–1.5 cm, respiration management in 18 patients</td>
<td>NTCP lung &lt;25%, cord &lt;50 Gy</td>
<td>RTOG ≥G3 toxicity in &gt;20%, MTD 84 Gy, 9 patients ineligible for planned dose step because of lung dose limits</td>
</tr>
<tr>
<td>PTV = GTV + 1 cm or more</td>
<td>Lung V20 ≤30%, mean esophagus dose ≤35 Gy, esophagus V5 ≤30%, spinal cord &lt;45 Gy, heart ≤40 Gy, liver ≤30 Gy, half liver ≤35 Gy</td>
<td>MTD 74 Gy (2 Gy/Fx)</td>
</tr>
<tr>
<td>PTV = GTV + 1–2 cm, elective LN</td>
<td>Cord &lt;50 Gy, lung V20 ≤35%, bronchial plexus &lt;66 Gy</td>
<td>CTC Grade 3–4 nonhematological toxicity, G4 esophagus toxicity, MTD not reached</td>
</tr>
<tr>
<td>PTV = GTV + 1–2 cm, elective LN</td>
<td>Group I, Lung V20 &lt;25%; Group II, Lung V20 25–37%; Group III, Lung V20 &gt;37%, cord &lt;45 Gy</td>
<td>RTOG G3 pneumonitis in ≥20%, MTD not reached</td>
</tr>
</tbody>
</table>

Table 1. Dose-escalation trials in non–small-cell lung cancer (Continued)
limited to 30% of the lung volume (3). The total lung volume was defined as the combined volume of the right and left lungs after subtraction of the GTV. The esophagus and spinal cord were expanded isotropically by 5 mm to obtain the respective planning organ-at-risk volumes (PRVs). The maximum allowed dose to the esophagus PRV was 55 Gy ($V_{55}$) to 30% of the organ volume, and 45 Gy to the spinal cord PRV (1, 6). The heart dose was limited to 40 Gy ($V_{40}$) for 50% of the total heart volume (21).

The “optimal” plan was converted to a deliverable plan by using segmental, multileaf collimator (MLC) step-and-shoot leaf sequences generated for a 21EX linear accelerator (Varian Medical Systems, Palo Alto, CA). Minimum field size and minimum monitor units per segment were kept the same for all plans. Reoptimization of the segment weights was performed to reduce the difference between the “optimal” and deliverable plans. Modifications of the planning technique might have resulted in unacceptable plans becoming eligible for higher dose levels. Usually, however, lowering the dose to one organ compromises the dose to other structures. Moreover, it was not the aim of this study to investigate methods of plan optimization, but to demonstrate a systematic problem in dose-escalation studies that potentially results in clinically relevant selection biases for higher dose levels.

Based on the IMRT plans for the different dose levels, dose–volume histograms (DVHs) were calculated for the PTV and for all the above-mentioned normal tissue structures. According to the results of the DVH evaluations, patients were divided into two groups: those in whom normal tissue dose constraints were met at all dose levels (“eligible” Group E), and those who were not eligible for higher doses (“ineligible” Group I), even though the planning objectives had been met at lower dose levels. For the two groups, different parameters were compared at the 66-Gy dose level.

Clinical and dosimetric parameters

The clinical and dosimetric parameters evaluated were (a) magnitude of PTVs, (b) monitor units (MUs), (c) overall plan quality, and (d) DVHs for the tumor and organs at risk. The overall plan quality is summarized in a score represented by a single number for the specified dose, dose-volume, or dose–response objective and (d) DVHs for the tumor and organs at risk. The overall plan quality, the lower the quality score.

Estimated radiobiological outcome parameters

Tumor-control probability (TCP) values were calculated based on the method of Webb and Nahum (22), using an $\alpha$ value of 0.23 Gy$^{-1}$ to describe the mean radiosensitivity of the cell population, and a standard deviation (SD) $\sigma_\alpha$ of 0.05 Gy$^{-1}$. A clonogenic cell density of $10^5$/cm$^2$ was assumed. The TCP values were compared at the 66-Gy dose level. These parameters were determined by Webb (23) to provide the best fit for a clinical squamous-cell carcinoma (SCC) data set compiled by Brenner (24). Though the data compiled by Brenner (24) were for SCCs of the upper digestive and respiratory tracts, the TCP values derived here for NSCLC were reasonable across the range of volumes and doses investigated. Normal tissue complication probabilities (NTCPs) were calculated for the lung using the model of Lyman (25), with a median toxic dose ($TD_{50}$) of 30.8 Gy, a steepness $m$ of 0.37, and a volume component $n$ of 0.99 (26). To calculate the NTCP values for the esophagus, a TD$_{50}$ of 47 Gy, a steepness $m$ of 0.36, and an $n$ value of 0.69 were used (1).

A statistical analysis using Fisher’s exact test was performed to check the significance of a subgroup of two patients not reaching the higher dose levels. If patient eligibility were independent of dose levels, one would expect the eligibility number to stay the same. The null hypothesis, therefore, is that the proportion of eligible patients at the lowest dose who become ineligible at the highest dose is zero.

Wilcoxon rank tests and $t$ tests were used to determine the significance of differences in clinical, dosimetric, and radiobiological parameters, with $p < 0.05$ to reject the null hypothesis.

RESULTS

Altogether, 7 of 9 patients were eligible for all dose levels (Group E). Two patients did not meet the eligibility criteria at the 70-Gy and 74-Gy dose levels, respectively, because the allowed lung dose limitations had been exceeded (Group I). In the patient who failed to be eligible for dose levels of $\geq$70 Gy, the maximum allowed spinal-cord dose also had been exceeded at the 78-Gy dose level. The second patient was

![Fig. 1. Dose comparison for a large, central lung tumor, with (a) 66-Gy and (b) 78-Gy dose levels for the same patient who was not eligible for doses $\geq$74 Gy. The macroscopic tumor is colored red. The 20-Gy isodose is indicated by the yellow outlines. The green outlines correspond with the 66-Gy isodose (a) and the 78-Gy isodose (b). Note the larger 20-Gy isodose area with the higher total dose.](image-url)
ineligible for the 74-Gy and 78-Gy dose levels only because of lung dose limits (Fig. 1). One of the 2 patients in Group I had a central left upper-lobe tumor with mediastinal lymph nodes; the second had a peripheral right lower-lobe tumor without enlarged lymph nodes. The fact that only 7 of 9 patients were eligible for all dose levels was statistically significant according to Fisher’s exact test (95% confidence interval, 0.075–0.6, significantly different from zero).

Clinical and dosimetric parameters

The average PTV volume of Group E was 132 cm$^3$, compared with 405 cm$^3$ for Group I ($p = 0.24$). Consequently, the average MUs per fraction for Group E were lower than for Group I (448 vs. 802, $p = 0.0008$). The overall plan quality score was better for Group E than for Group I (0.012 vs. 0.068; $p = 0.12$). The average dose to 95% of the PTV was higher in Group E than in Group I (65.25 Gy; range, 64.64–65.52 Gy; SD, 0.33 Gy) vs. 63.32 Gy; range, 63.05–63.58 Gy; SD, 0.37 Gy; $p = 0.047$), indicating that reaching the PTV objective was more difficult for Group I. The results from the evaluation of DVHs for normal tissue are given in Table 2. The difference in lung V$_{20}$ values was highly significant between Groups E and I ($p = 0.0002$), with patients in Group I on average having a 12% higher volume receiving doses $\geq 20$ Gy. Figure 2 gives an example of the lung DVHs of 2 patients: one was eligible for all dose levels, while the other was ineligible at the 74-Gy and 78-Gy dose levels.

Estimated radiobiological outcome parameters

The average TCP values were 33.1% (SD, 7.5%) for Group E, and 22.7% (SD, 9.2%) for Group I ($p = 0.10$). The average NTCP values for the lung were 2.1% (SD, 0.8%) for Group E, and 9.7% (SD, 1.4%) for Group I ($p = 0.028$). For the esophagus, NTCP values of 3.4% (SD, 0.9%) for Group E, and 12.1% (SD, 6.5%) for Group I, were calculated ($p = 0.68$).

DISCUSSION

Eligibility criteria in radiotherapy dose-escalation trials regularly include dose limitations to normal tissue. While

| Table 2. Dose–volume evaluation for Groups E and I at the 66-Gy dose level |
|---------------------------|---------------------------|---------------------------|---------------------------|
|                         | Group E |               | Group I |               | $p$ value |
|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Lung V$_{20}$ Gy, (%)     | 15.3      | 5.0           | 27.2     | 0.7           | 0.00002 |
| Esophagus V$_{55}$ Gy (%) | 7.7       | 5.4           | 16.9     | 12.7          | 0.23    |
| Heart V$_{40}$ Gy (%)     | 2.7       | 2.6           | 18.1     | 11.7          | 0.16    |
| Spinal cord D$_{0.1%}$ (Gy)| 35.2      | 11.1          | 41.2     | 5.0           | 0.17    |

Abbreviations: D$_x$ = minimum dose received by volume $x$ of a specified organ; V$_x$ = volume of a specified organ receiving dose of $x$ Gy or more.

normal tissue constraints generally remain constant in dose-escalation trials, the tumor dose is increased for each dose level. Because tumor dose and normal tissue constraints are competing demands, a point will be reached at which tumor dose cannot be increased without exceeding normal tissue constraints. If this point is reached within the dose spectrum of a dose-escalation trial, a selection bias might be introduced.

In this virtual dose-escalation trial in patients with NSCLC, we found that some patients were not eligible for higher dose levels, even though they had met the eligibility criteria at lower dose levels. As a result, the estimated outcome between patients who had been eligible for higher doses and those who had met the selection criteria only for the lower doses was significantly different. This might have led to an inadequate determination of maximum tolerated doses (MTDs). Though the sample size of this study was
relatively small, the sample was large enough to demonstrate the presence of a selection bias with statistical significance, emphasizing the magnitude of the effect.

In a real trial, patients are enrolled in a study without knowing their definitive dose level. When it is seen later that patients are not eligible for higher dose levels because of normal tissue dose limitations, they are either excluded from the study or treated at lower dose levels. At the same time, patients with more favorable dosimetric parameters are recruited to higher dose levels. In general, planning objectives are met more easily for small tumors. Tumor size is considered a predictive factor for the treatment outcome in patients with NSCLC tumors (9, 27–30), and might therefore confound results in dose-escalation trials. Kong et al. (8) reported that in their trial, patients receiving lower doses generally had the larger tumors. However, they found that radiation dose was a stronger predictor than tumor volume generally had the larger tumors. However, they found that radiation dose was a stronger predictor than tumor volume.

Rosenzweig et al. (10) suggested subdividing tumors into small, intermediate, and large, to better identify the MTD that is valid for a respective tumor size. In the Radiation Therapy Oncology Group (RTOG) 9311 protocol, patients were stratified according to lung V20. Patients with a V20 of <25% had smaller tumors, compared with patients with a V20 between 25–36%, and were therefore eligible for higher dose levels (2).

The PTV size varies considerably, depending on the use of respiration management or other image-guidance strategies. In this study, isotropic margins were used to imitate the concepts of current dose-escalation studies, with the treatment scenario being respiratory-gated IMRT. If no respiration management or other image-guidance strategies had been assumed, a change of safety margins would have been appropriate, depending on the methodology and patient-specific findings. Changing the safety margin might lead to a different portion of patients becoming eligible for higher dose levels. Motion-encompassing treatment without respiration management, including the target volumes from each phase of the 4D CT scan covering the whole range of tumor motion, will lead to an expansion of PTVs and a lower percentage of patients eligible for all dose levels. For example, creating a PTV that comprised the CTVs of all respiratory phases enlarged the PTVs of the two Group I patients by a factor of 1.26. This resulted in an ineligibility of both patients even at the lowest dose level of this study. Reducing the CTV to PTV safety margin from 8 to 5 mm by using improved image guidance still left 1 of 2 patients ineligible for the 66-Gy dose level. Margin definition, respiration management, and image guidance are therefore important parameters to consider during the design and interpretation of dose-escalation studies.

In two dose-escalation trials, the authors mentioned that patients who were not eligible for higher dose levels were treated with lower doses (7, 10). In one report, the lung dose limits were exceeded in 9 of 104 patients (10). In the other study, which included 55 patients, some patients did not meet the dose constraint thresholds of the protocol, 1 patient for spinal cord, 6 patients for esophagus (7). In a third study, 3 patients were either ineligible for a certain dose level, or eligibility could not be determined (2). The exact background for ineligibility was not mentioned.

As we showed in the present study, treating patients who are ineligible for higher dose levels (e.g., because of unfavorable tumor geometry, larger tumor size, or mediastinal lymph node involvement) (8) with lower doses introduces a statistically significant selection bias. With ranges between 9 and nearly 50 Gy for minimum and maximum dose levels in published dose-escalation trials, this phenomenon is probably observed more frequently.

The use of constant normal organ dose limitations as part of the eligibility criteria already preselects more favorable patients for study enrollment. While this does not introduce a bias per se, it is important to remember that MTDs are only valid for specified patient and treatment criteria. As an example, the 9 patients described in this study are a subgroup of a 13-patient cohort, in which 4 patients were not even eligible for the 66-Gy dose level. A comparison between the 9 eligible and the 4 ineligible patients revealed higher NTCP values for the lung and esophagus (p = 0.08 and 0.06, respectively), and significantly lower TCP values (p = 0.02) for the 4 ineligible patients.

To address the higher risk of pneumonitis for higher lung doses, some studies stratified patients according to calculated risks, and allowed higher dose levels only for low-risk patients. Stratifications were performed according to the magnitude of lung V20, the NTCP values for the lung, or the relative mean lung dose levels (2, 7, 8, 10, 12). Although this differentiation is very helpful for delivering higher dose levels safely and for defining the MTD for subgroups of patients, doses might still vary up to >30 Gy (7) within each risk group, resulting in a still significant potential for selection biases.

The interpretation and comparison of dose-escalation studies is certainly a challenging task. Variations in clinical eligibility criteria (such as tumor stage, lung-function parameters, and the application and timing of chemotherapy) limit the direct comparability of studies. In addition, differences in technical parameters, such as treatment technique (IMRT vs. three-dimensional conformal RT), choice of safety margins and target volumes (e.g., with or without elective lymph nodes), and the use of different fractionation schemes (such as accelerated, normofractionated, or hypofractionated treatments), obviate a direct comparison (Table 1). As a result, the level of MTDs in eligible patients varies from 74 Gy to >100 Gy in studies in which the MTD has not even been reached. Moreover, different studies use different toxicity classification systems to define the MTD. Dosimetric eligibility criteria most commonly use lung dose as a limit; however, some studies also use spinal cord, esophagus, or heart dose to identify eligible patients. The selection bias observed in the present study adds to these factors, influencing the interpretation of dose-escalation studies, and as we have shown, is of clinical relevance.

How can this selection bias be reduced or eliminated? One straightforward method is to select only those patients
for a dose-escalation trial for whom normal tissue tolerances meet the eligibility criteria at the highest dose level of the study, irrespective of the dose level actually planned for each patient. This will guarantee that all patients within a trial or stratification group have similar planning parameters and comparable outcome estimates.

In conclusion, the present use of constant normal organ dose limitations is assumed to introduce a clinically relevant selection bias in dose-escalation trials that might result in a generalization of treatment results that is applicable only to a subgroup of the patient population. This selection bias can be avoided by including only patients who qualify for all dose levels.

REFERENCES