Improvement of breast cancer irradiation techniques
Hurkmans, C.W.

Citation for published version (APA):
Chapter 2

Variability in target volume delineation on CT scans of the breast

Coen W. Hurkmans
Jacques H. Borger
Bradley R. Pieters
Nicola S. Russell
Edwin P.M. Jansen
Ben J. Mijnheer

The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital
Department of Radiotherapy, Plesmanlaan 121, 1066 CX Amsterdam
The Netherlands

Chapter 2

Abstract

Purpose. To determine the intra- and interobserver variation in delineation of the target volume of breast tumors on CT scans in order to perform conformal radiotherapy.

Methods and materials. The clinical target volume (CTV) of the breast was delineated in CT slices by four radiation oncologists on our clinically used delineation system. The palpable glandular breast tissue was marked with a lead wire on 6 patients before CT scanning, while four patients were scanned without a lead wire. The CTV was drawn three times by each observer at separate occasions. Planning target volumes (PTVs) were constructed by expanding the CTV by 7 mm in each direction, except towards the skin. The deviation in the PTV extent from the average extent was quantified in each orthogonal direction for each patient to find a possible directional dependence in the observer variations. In addition, the standard deviation of the intra- and interobserver variation in the PTV volume was quantified. For each patient the common volume delineated by all observers and the smallest volume encompassing all PTVs were also calculated.

Results. The patient averaged deviations in PTV extent were larger in the posterior (42 mm), cranial (28 mm) and medial (24 mm) directions than in the anterior (6 mm), caudal (15 mm) and lateral (8 mm) directions. The mean intraobserver variation in volume percentage (5.5%, 1 SD) was much smaller than the interobserver variation (17.5 %, 1 SD). The average ratio between the common and encompassing volume for the four observers separately was 0.82, 0.74, 0.82 and 0.80. A much lower combined average ratio of 0.43 was found due to the large interobserver variations. The intraobserver variation was decreased by a factor 4 on scans made with a lead wire compared to scans made without a lead wire for the observer that placed the lead wire. For the other observers no improvement was seen. Based on these results an improved delineation protocol was designed.

Conclusions. Intra- and especially interobserver variation in the delineation of breast target volume on CT scans can be rather large. A detailed delineation protocol, making use of CT scans with lead wires placed on the skin around the palpable breast by the delineating observer reduces the intraobserver variation. Better imaging techniques and pathology studies relating glandular breast tissue with imaging may be needed to provide more information on the extent of the clinical target volume in order to reduce the interobserver variation.
Introduction

Radiation therapy plays an important role in the management of patients with breast cancer. Conventionally, postoperative radiotherapy of the breast is delivered with rectangular tangential fields. The position of these fields is usually determined at a simulator and encompasses the palpable glandular breast tissue. The optimal gantry angle and collimator angle of these fields depends on the medial and lateral extension of the palpable glandular breast tissue and the curvature of the chest wall. With this technique there are appreciable dose variations within the treatment portals (1,2), while the dose to the lung and heart can still be high (3-6). Intensity modulation treatment techniques have been developed in order to reduce these dose variations. Some of these techniques use tangential fields with a straight dorsal field edge (7-9). However, better conformation of the treatment fields to the target volume is needed to obtain a further reduction of the dose to the organs at risk and to prevent underdosage of part of the target volume.

A prerequisite for conformal radiotherapy is accurate knowledge of the position and shape of the target volume. Therefore, it is important to quantify the variation in the delineation of the target volume. Valdagni et al. found large differences between clinical target volume (CTV) localization using standard anatomical borders, palpation and ultrasound (10). Bentel et al. showed that field borders based on surface anatomy are often significantly different from field borders based on CT-information (11). Therefore, treatment techniques using conformal fields (12-16) should be implemented with care, taking into consideration the uncertainty in the delineation of the breast target volume.

Systematic quantification of intra- and interobserver variation in the delineation of CTVs has been performed for a number of treatment sites, e.g., prostate, lung and brain tumors (17-22). Overall observer variations were relatively small for prostate tumors (approximately 3 mm, 1 SD), compared to lung tumors, where differences of several centimeters have been observed. Many studies show that the interobserver variation is significantly larger than the intraobserver variation. However, such data are not available for the breast.

The aim of this study is therefore to determine the intra- and interobserver variation in the determination of the target volume of the breast on CT scans.
Chapter 2

Methods and materials

Patient data

The data of 7 patients with breast cancer (of which five left-sided and two right-sided) were used in this study. CT scanning was performed with the patient in supine position, arms lying along the side of the body. This procedure is not used clinically, as our treatment fields are routinely defined using a simulator. The slice separation and slice thickness was 5 mm. The scans included the whole breast with at least a 30 mm margin in the cranial and caudal direction. For 6 patients, the extent of the palpable glandular breast tissue was marked by one of the observers with a small lead wire prior to the scan. For three of these patients, a scan was also made without the lead wire. For patient 7, only a scan without a lead wire was made. Thus, ten different CT-scan sets were available for this study. The patient age varied between 46 and 79 years at the time of the CT-scan.

Target volume delineation

The CTV was outlined on axial CT slices using a delineation protocol, of which a brief outline will be given. A standard window level (0) and width (500) that was considered optimal for visibility of the glandular breast tissue was used to display the images. If no lead wire was present, the delineation of the CTV was based on the visible breast parenchyma. If a lead wire was present, the position of the wire could be used as an aid in the delineation of the CTV. It was clearly stated in the protocol that, based on the CT data, the CTV contours could still extend outside the region defined by the lead wire, or be smaller than this region. After delineation, a 3D view was made. Thereafter, the delineations could be edited based on this 3D view. Then, after editing, a new 3D view had to be made to check whether the target surface was smoothed properly. If this surface was not smooth, consecutive re-editing could be performed until the observer was satisfied with the results. The surface of the delineated CTV should be smooth, as it was assumed that the CTV is smooth in reality as well. The CTV was assumed to start 5 mm below the skin. Four radiation oncologists delineated the CTV three times for each patient. The CTV was delineated in all slices, except in the middle part of the breast around the nipple, where every other slice could be skipped if no substantial contour changes were observed. A minimum of one week was required between consecutive delineations of a particular patient. None of the observers had knowledge of the CTVs outlined by the other observers.

A planning target volume (PTV) was constructed by expanding the CTV
Breast target volume delineation

7 mm. This margin was estimated to be adequate to account for the uncertainty in the patient set-up and CTV delineation. The expansion from CTV to PTV was not performed towards the skin. The data analysis was performed for the PTV, because it represents the volume used to define the shape of the beams. This volume is the volume used to prescribe and report dose and therefore the volume on which a treatment plan and its evaluation are based (23,24). Hence, it can currently be considered the most relevant volume for quantification of delineation variations.

Data analysis

The most medial, lateral, anterior, posterior, cranial and caudal extensions of the PTVs were determined. The extensions were determined by automated calculation of the coordinates of a box just encompassing the planning target volume. Thereafter, the deviations of all extensions from the average extension were calculated to find a possible directional dependence in the observer variations.

Considering the group of four observers, with three delineations per observer for each of the ten CT-scan sets, the average volume $V$ was defined as the average of all $N$ volumes outlined in one patient:

$$V = \frac{1}{N} \sum_{o=1}^{O} \sum_{d=1}^{D} V_{od} = \frac{1}{12} \sum_{o=1}^{4} V_{o}$$

where $O$ denotes the number of observers and $D$ the number of delineations of one observer in one CT-scan set. $V_{od}$ denotes the volume of a delineation of an observer in the specific CT-scan set and $V_o$ the average volume of the three delineations of one observer. The standard deviation of the intraobserver variation around this mean is given by:

$$\sigma_{o} = \sqrt{\frac{1}{D-1} \sum_{d=1}^{D} \left(V_{od} - V_{o}\right)^2} = \sqrt{\frac{1}{2} \sum_{d=1}^{3} \left(V_{od} - V_{o}\right)^2}.$$

The observer averaged intraobserver variation was calculated according to:

$$\sigma = \sqrt{\frac{1}{N-O} \sum_{o=1}^{O} (D-1)\sigma_{o}^2} = \sqrt{\frac{1}{4} \sum_{o=1}^{4} \sigma_{o}^2}.$$
The standard deviation of the interobserver variation is given by:

\[
\Sigma = \sqrt{\frac{O}{N(O-1)} \sum_{o=1}^{O} D(V_o - V)^2} = \sqrt{\frac{1}{3} \sum_{o=1}^{4} (V_o - V)^2}.
\]

Because of the limited sample size when computing \(V_o\), the standard deviation of the interobserver variation will contain an intraobserver component. An estimate of the interobserver variation, correcting for this effect, is given by:

\[
\Sigma^* = \sqrt{\Sigma^2 - \frac{\sigma^2}{N/O}} = \sqrt{\Sigma^2 - \frac{\sigma^2}{3}}.
\]

Furthermore, the encompassing volume for one observer and for all four observers was defined as the volume encompassing the PTVs of one or all four observers in one patient, respectively. The common volume for one observer and for all four observers was defined as the volume common to the PTVs of one or all four observers, respectively. Thereafter, the ratio between the common volume and encompassing volume was calculated for each observer individually and for all four observers combined. This ratio can vary between 0, which means all delineations define a completely different volume, and 1, which means all volumes are completely identical without any intra- or interobserver variation. Therefore, this ratio is not only a measure for the differences in volume between delineations, but also a measure for the difference in shape and place of the delineations. A comparison was also made of variations in delineations between patients marked with a lead wire (scan sets 1L to 6L) and patients without lead wire marking (scan sets 3,4,6 and 7).

**Table 1.** Maximum deviations from the average extent of the PTV in three orthogonal directions (combined intra- and interobserver differences).

<table>
<thead>
<tr>
<th>Observer</th>
<th>Anterior</th>
<th>Posterior</th>
<th>Caudal</th>
<th>Cranial</th>
<th>Lateral</th>
<th>Medial</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.4</td>
<td>4.8</td>
<td>2.0</td>
<td>1.5</td>
<td>1.5</td>
<td>3.7</td>
</tr>
<tr>
<td>B</td>
<td>2.1</td>
<td>8.0</td>
<td>6.5</td>
<td>11.0</td>
<td>2.5</td>
<td>12.2</td>
</tr>
<tr>
<td>C</td>
<td>1.7</td>
<td>8.4</td>
<td>3.5</td>
<td>5.0</td>
<td>2.9</td>
<td>8.1</td>
</tr>
<tr>
<td>D</td>
<td>1.0</td>
<td>12.0</td>
<td>3.0</td>
<td>12.5</td>
<td>4.4</td>
<td>8.5</td>
</tr>
<tr>
<td>Average</td>
<td>1.6</td>
<td>8.3</td>
<td>3.8</td>
<td>7.5</td>
<td>2.8</td>
<td>8.1</td>
</tr>
<tr>
<td>Combined†</td>
<td>5.5</td>
<td>42.1</td>
<td>14.5</td>
<td>27.5</td>
<td>8.4</td>
<td>24.5</td>
</tr>
</tbody>
</table>

*Average, intraobserver variation  
† Combined, intra- and interobserver variation*
Results

The maximum difference in the extension of all 12 PTVs per patient (i.e., including the results of all observers) was largest in the posterior direction (42 mm, Figure 1). This is mainly caused by the difference between the observers in the delineation of the posterior part of the target volume in the CT slice that includes the nipple (Figure 2). The extent of the target volume in the anterior and lateral direction is determined by the patient outline. Therefore, the differences in the anterior (6 mm) and lateral (8 mm) directions are small. As can be seen in Figure 1, the differences also vary between patients. The larger lateral difference for patient 3 was caused by differences in the delineation of the target volume close to the axilla for that specific patient. As can be seen in Figure 1, the variation in the cranial and caudal directions is discrete with a 5 mm step size, which is equal to the CT slice separation.

Figure 1. Maximum differences in the extent of the PTVs delineated per patient, averaged over all patients.
Chapter 2

The maximum differences were also calculated separately for each patient for all observers separately, based on three PTVs per patient, and combined for all four observers, based on 12 PTVs per patient. Thereafter, the average over all patients for each observer (intraobserver variation) and for all four observers combined (combined intra- and interobserver variation) were calculated (Table 1).

Figure 2. Breast cross sections of 12 PTV delineations in an axial slice through the middle part of the breast.

The combined differences were a factor of 3-5 larger than the intraobserver differences in all directions, indicating that the interobserver variation in extension is of more importance than the intraobserver variation.

The patient averaged PTV ranged from 280 cm$^3$ for patient 4 to 2228 cm$^3$ for patient 6 (Table 2). The PTV defined by the more experienced observers A and B was for all 10 CT-scan sets smaller, on average 20% and 33%, respectively, than the average volume defined by the two other observers. The intraobserver variation ranged from 3% to 7% of the average volume (1 SD). The interobserver variation corrected for the small sample size, ranged between 11% and 27% (1 SD). On average, the interobserver variation was approximately three times larger than the intraobserver variation.
Table 2. Average volume and ratios between common and encompassing volume for the four observers separately.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Observer</th>
<th>V_1</th>
<th>R_1</th>
<th>V_1</th>
<th>R_1</th>
<th>V_1</th>
<th>R_1</th>
<th>V_1</th>
<th>R_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1L</td>
<td>A</td>
<td>668</td>
<td>0.90</td>
<td>555</td>
<td>0.78</td>
<td>731</td>
<td>0.87</td>
<td>752</td>
<td>0.77</td>
</tr>
<tr>
<td>2L</td>
<td>B</td>
<td>1106</td>
<td>0.81</td>
<td>997</td>
<td>0.84</td>
<td>1288</td>
<td>0.92</td>
<td>1302</td>
<td>0.85</td>
</tr>
<tr>
<td>3L</td>
<td>C</td>
<td>369</td>
<td>0.86</td>
<td>257</td>
<td>0.74</td>
<td>422</td>
<td>0.78</td>
<td>374</td>
<td>0.79</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>280</td>
<td>0.73</td>
<td>287</td>
<td>0.72</td>
<td>419</td>
<td>0.77</td>
<td>399</td>
<td>0.74</td>
</tr>
<tr>
<td>4L</td>
<td>E</td>
<td>274</td>
<td>0.84</td>
<td>236</td>
<td>0.72</td>
<td>318</td>
<td>0.81</td>
<td>324</td>
<td>0.78</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>274</td>
<td>0.72</td>
<td>223</td>
<td>0.67</td>
<td>295</td>
<td>0.79</td>
<td>328</td>
<td>0.78</td>
</tr>
<tr>
<td>5L</td>
<td>G</td>
<td>643</td>
<td>0.88</td>
<td>522</td>
<td>0.74</td>
<td>789</td>
<td>0.82</td>
<td>869</td>
<td>0.74</td>
</tr>
<tr>
<td>6L</td>
<td>H</td>
<td>1835</td>
<td>0.88</td>
<td>1560</td>
<td>0.72</td>
<td>2418</td>
<td>0.86</td>
<td>2426</td>
<td>0.86</td>
</tr>
<tr>
<td>6</td>
<td>I</td>
<td>1974</td>
<td>0.80</td>
<td>1689</td>
<td>0.76</td>
<td>2652</td>
<td>0.84</td>
<td>2597</td>
<td>0.86</td>
</tr>
<tr>
<td>7</td>
<td>J</td>
<td>1031</td>
<td>0.83</td>
<td>706</td>
<td>0.68</td>
<td>1289</td>
<td>0.69</td>
<td>1155</td>
<td>0.83</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.82</td>
<td>0.74</td>
<td>0.82</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* V_1, average volume (cm³) based on the delineations of one observer
† R_1, ratio between common and encompassing volume based on the delineations of one observer

The ratio between the common and encompassing volume for the four observers separately ranged from 0.67 to 0.92. The average ratio for the observers was 0.82, 0.74, 0.82 and 0.80, implying that the intraobserver variation was approximately equal for all observers. A much lower combined average ratio of 0.43 was found due to the large interobserver variations (Table 3).

The effect of the presence of a lead wire around the palpable glandular breast tissue on the intra- and interobserver variation was quantified by comparing the delineation results in scan sets 3L, 4L and 6L to scan sets 3, 4 and 6, respectively. The intraobserver variation, expressed by the ratio between encompassing and common volume, decreased for all observers when a lead wire was present. The improvement was most pronounced for observer A, who placed the lead wire. The ratio improved from 0.73 to 0.86 for patient 3, from 0.72 to 0.84 for patient 4 and from 0.80 to 0.88 for patient 6. Only in one case (observer B, patient 6) a decrease of the ratio from 0.76 to 0.72 was found. The intraobserver variation in volume σ (% volume, 1 SD) decreased considerably from 10% to 2% for patient 3, from 3% to 1% for patient 4 and from 5% to 1% for patient 6 for observer A that placed the lead wire. For the other observers almost no improvement was seen. The improvement was most pronounced for observer A, because, if a lead wire was present, he delineated a volume that was more similar to the volume marked by the lead wire than the other observers. The interobserver

31
variation in volume did not improve by the placement of a lead wire.

The differences in shape of the PTV were large as can be seen in the example given in Figure 3. The more experienced observers delineated a more or less round target volume as seen from anterior, while the other observers delineated a somewhat cylindrically shaped volume. This difference might be due to the larger doubt non-experienced observers have in the circumscription of the microscopic extension of the target volume. In order to prevent missing a part of the clinical target volume these observers are inclined to delineate a larger target volume compared with more experienced observers. Starting with a large contour at the position of the nipple, these observers are not inclined to decrease the contour size considerably in the cranial and caudal slices.

Table 3.  Average volume and ratios between common and encompassing volume for all four observers combined.

<table>
<thead>
<tr>
<th>Patient</th>
<th>V* (cm$^3$)</th>
<th>R †</th>
<th>σ ‡</th>
<th>Σ§</th>
</tr>
</thead>
<tbody>
<tr>
<td>1L</td>
<td>677</td>
<td>0.51</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>2L</td>
<td>1173</td>
<td>0.58</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>3L</td>
<td>355</td>
<td>0.41</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>346</td>
<td>0.39</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>4L</td>
<td>288</td>
<td>0.43</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>280</td>
<td>0.39</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>5L</td>
<td>706</td>
<td>0.39</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>6L</td>
<td>2060</td>
<td>0.46</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>2228</td>
<td>0.46</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>1045</td>
<td>0.33</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>Mean</td>
<td>0.43</td>
<td>5.5</td>
<td>17.5</td>
<td></td>
</tr>
</tbody>
</table>

* V, average volume (cm$^3$) based on the delineations of all four observers
† R, ratio based on the delineations of all four observers
‡ σ, intraobserver variation (%volume, 1 SD)
§ Σ, corrected interobserver variation (%volume, 1 SD)
Breast target volume delineation

Figure 3. Shape of a PTV of patient 7 as delineated by observers B and D.

Discussion

The large differences found in the cranial and posterior direction were caused by differences in the interpretation of extent of the CTV in the direction of the axilla and dorsally along the thoracic wall. The difficulty in interpretation of the extent of the CTV is caused by the lack of a clear border in the CT images between the breast parenchyma and fatty tissue. The breast parenchyma is slightly better visible in MR images. However, the gradual transition between the location of the glandular breast tissue and the fat tissue will also in MR images lead to an unclear border between these tissues.

The average range in the maximum cranial and caudal deviations is much higher than the discrete step size of 5 mm of the CT slice separation. It is therefore unlikely that a smaller CT slice separation will reduce these deviations appreciably.

There was not always a gradual transition in the delineations between adjacent axial CT slices. This was caused by the inability of our current software to view a complete 3D representation of the target during the delineation process. In general, editing was done only once, because the
method of displaying the 3D view and re-editing is very labor-intensive. On average, delineation of one target volume using this protocol took approximately 30 minutes.

**Table 4.** Ratios between common and encompassing volume for the four observers and the results obtained using the improved protocol.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Observer</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Average</th>
<th>Improved protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>3L</td>
<td>A</td>
<td>0.86</td>
<td>0.74</td>
<td>0.78</td>
<td>0.79</td>
<td>0.79</td>
<td>0.86</td>
</tr>
<tr>
<td>4L</td>
<td>B</td>
<td>0.84</td>
<td>0.72</td>
<td>0.81</td>
<td>0.78</td>
<td>0.79</td>
<td>0.88</td>
</tr>
<tr>
<td>6L</td>
<td>C</td>
<td>0.88</td>
<td>0.72</td>
<td>0.86</td>
<td>0.86</td>
<td>0.83</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Using the results from this study, we improved our delineation protocol. Firstly, a radiation oncologist should always place a lead wire around the palpable glandular breast tissue. The delineation of the CTV should be performed by the radiation oncologist who placed the lead wire, or by a radiation oncologist who was also present at the time the lead wire was placed. Secondly, the CTV should be located inside the volume defined by the lead wire, unless glandular breast tissue is still visible outside this volume. Thirdly, delineation has to be performed using more advanced software, capable of displaying the delineation simultaneously in axial, sagittal and coronal projections and in a 3D view. Also, editing of the delineations should be more user friendly than in our current software. For example, the change in shape of the volume should directly be seen in 3D while editing, without leaving the editing menu. The preliminary results presented in Table 4 suggest that the new guidelines result in a faster and more consistent delineation of the target volume. In order to prevent streaking artifacts of the wire in the CT images, we tested if an electricity wire of 1.1 mm in diameter with a copper core of 0.5 mm would also be visible in a CT image. It appeared to be clearly visible without causing any streaking artifacts.

Often, it is assumed that the CTV includes the complete palpable breast. The results presented here are based on this assumption. However, it could be very meaningful to study the location of the glandular breast tissue in lumpectomy and mastectomy samples. This could lead to a better discrimination between the part of the palpable breast that contains glandular breast tissue and the part that does not contain glandular breast tissue and could be excluded from the CTV.

The intra- and interobserver variations in the delineation of breast target volume on CT scans can be very large. A detailed delineation protocol, making use of CT scans with lead wires placed on the skin around the palpable breast, combined with delineation software with full 3D viewing capacities can result in a small intraobserver variation. Studies relating the
Breast target volume delineation

position of the glandular breast tissue in the breast using lumpectomy or mastectomy samples to breast imaging techniques might provide us more information on the extent of the CTV. This information could lead to a reduction of the interobserver variations in the delineation of the CTV.

References


