Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them?

Barbara Segedin¹, Primoz Petric²

¹ Department of Radiation Oncology, Institute of Oncology Ljubljana, Slovenia
² Department of Radiation Oncology, National Centre for Cancer Care and Research, Doha, Qatar

Background. Modern radiotherapy techniques enable delivery of high doses to the target volume without escalating dose to organs at risk, offering the possibility of better local control while preserving good quality of life. Uncertainties in target volume delineation have been demonstrated for most tumour sites, and various studies indicate that inconsistencies in target volume delineation may be larger than errors in all other steps of the treatment planning and delivery process. The aim of this paper is to summarize the degree of delineation uncertainties for different tumour sites reported in the literature and review the effect of strategies to minimize them.

Conclusions. Our review confirmed that interobserver variability in target volume contouring represents the largest uncertainty in the process for most tumour sites, potentially resulting in a systematic error in dose delivery, which could influence local control in individual patients. For most tumour sites the optimal combination of imaging modalities for target delineation still needs to be determined. Strict use of delineation guidelines and protocols is advisable both in everyday clinical practice and in clinical studies to diminish interobserver variability. Continuing medical education of radiation oncologists cannot be overemphasized, intensive formal training on interpretation of sectional imaging should be included in the program for radiation oncology residents.

Key words: target volume; interobserver variability; delineation uncertainties; imaging; training
In addition, such uncertainties can undermine meaningful comparison of treatments within and between institutions and interpretation of clinical studies.

Uncertainties in target volume delineation have been demonstrated for most tumour sites, and various studies indicate that inconsistencies in target volume delineation may be larger than errors in all other steps of the treatment planning and delivery process.\textsuperscript{7-22}

The aim of this paper is to summarize the degree of delineation uncertainties for different tumour sites reported in the literature and review the effect of strategies to minimize them.

**Magnitude of uncertainties**

Direct comparison of published data is difficult, since a variety of methods is used to quantify interobserver variability. Most papers report parameters describing the distribution of delineated volumes including mean, range, standard deviation (SD), the ratio of the largest and the smallest delineated volume ($V_{\text{max}}/V_{\text{min}}$), coefficient of variation (COV) etc. Also commonly used are different concordance measures such as conformity, concordance or similarity index (CI, SI – ratio between common and encompassing volume), Dice-Jaccard coefficient (DJC), percent overlap, ratio of encompassing and common volume ($1/CI$), geographical miss index and mean discordance index or statistical measures of agreement i.e. kappa ($\kappa$) - statistics.\textsuperscript{23} Less commonly, methods for local interobserver variation assessment are used i.e. local standard deviation (SD), inter-delineation distance or radial line measurement variation, all expressed in mm.\textsuperscript{24-27}

A wide range of interobserver variability is observed for various tumour sites, the largest variation being reported for target volume delineation in oesophageal, head and neck, Hodgkin’s lymphoma and sarcoma, where the $V_{\text{max}}/V_{\text{min}}$ ratios are 6, 18.3, > 7, 15 and > 8, respectively.\textsuperscript{3,18,28-30}

**Gastrointestinal tumours**

In rectal cancer, reported conformity indices are from 0.29 to 0.98 for clinical target volume (CTV) and from 0.26 to 0.81 for primary tumour gross target volume (GTV), depending on the use of consensus guidelines and chosen imaging modality.\textsuperscript{19,31-33} The ratio of $V_{\text{max}}/V_{\text{min}}$ is 1.93 – 2.65 for GTV and 1.75 – 4.71 for CTV depending on imaging modality used for treatment planning.\textsuperscript{19} Interobserver variability in CTV and planning target volume (PTV) delineation for gastric cancer was assessed as a part of the CRITICS trial. Despite delineation atlas provided for participants $V_{\text{max}}/V_{\text{min}}$ ratio was 3.4 for CTV and 2.6 for PTV and the authors speculated the reason was unfamiliarity with target volumes in the upper abdomen.\textsuperscript{8} For oesophageal cancer, median Jaccard conformity index for GTV was 0.69 in a study by Gwynne \textit{et al.}\textsuperscript{34}, with the highest observer agreement in the middle section of the GTV, which is a marked improvement compared to results reported by Tai \textit{et al.}\textsuperscript{18} at the start of 3D planning era, when $V_{\text{max}}/V_{\text{min}}$ ratio was up to 6.

**Cervix cancer**

Considerable interobserver variability was described by Weiss \textit{et al.}\textsuperscript{10} in CTV for cervix carcinoma, with the ratio of common to encompassing volume from 0.11 to 0.57 and $V_{\text{max}}/V_{\text{min}}$ ratio 1.3 – 4.9. The main reason for large variability was wide variation in caudal and cranial CTV borders, resulting from varying inclusion of specific nodal regions (para-aortic, iliac and inguinal) by the observers. In a study of cervix cancer IGABT, CI was 0.6 – 0.8 for high risk CTV (HR CTV) and 0.6 – 0.7 for GTV and intermediate risk CTV (IR CTV), demonstrating a relatively good interobserver agreement considering that CI is sensitive to volume size and volumes in brachytherapy tend to be much smaller than in EBRT. Mean inter-delineation distance was 4.2 mm, 3.8 mm and 5.2 mm for GTV, HR CTV and IR CTV, respectively.\textsuperscript{25,35}

**Head and neck tumours**

Interobserver variability in CTV delineation in oropharyngeal cancer (tonsillar tumour) is one of the largest described in the literature. With the primary GTV already provided, $V_{\text{max}}/V_{\text{min}}$ ratio for CTV reported by Hong \textit{et al.}\textsuperscript{30} was 18.3. Recommended PTV expansion from the contoured CTV also varied considerably in different institutions (mean 4.11 mm, range 0 – 15 mm). Smaller but still significant variability was reported by Thiagarajan \textit{et al.}\textsuperscript{8} for oropharyngeal primary tumour GTV with CI 0.54 – 0.62, depending on imaging modality. Agreement on nodal GTV was higher with CI > 0.75 for all imaging modalities. For nasopharyngeal carcinoma local SD was 3.3 – 4.4 mm for CTV (visible tumour + potential microscopic extension) and 4.9 – 5.9 mm for elective CTV (CTV + 1 cm margin and the entire nasopharynx), depending on imaging modality.\textsuperscript{24}
Lung tumours

In lung cancer the range of reported interobserver variability is quite large with $V_{\text{max}}/V_{\text{min}}$ ratio from 1.8 to 2.3 for primary GTV alone and from 5.2 to $>7$ for primary and nodal GTV. Reported conformity indices range from 0.04 to 0.70 for the same target volumes, depending on imaging modality, with some authors describing cases where there was no common volume for all observers.\(^3,14,15,27,36\)
Like in cervix cancer the reason for large variability is inclusion of different nodal regions in the target volume. In a study by Van De Steene et al.\(^3\) the observers included only 63% of involved nodal regions in the target volume (generating 37% false negative nodes), on the other hand 22% of included nodal regions were considered false positive after a review. The authors suggested lack of knowledge being one of the main reasons for interobserver variability, beside problems of methodology (interpretation of GTV definition, drawing precision etc.) and difficulty in discriminating the tumour from surrounding pathological (\textit{i.e.} atelectasis, peritumoral reaction) and normal structures (\textit{i.e.} mediastinal vessels).

Other tumour sites

Interobserver variability for target delineation in brain tumours is similar to the one described for prostate with $V_{\text{max}}/V_{\text{min}}$ ratio from 1.3 to 2.8 and CI from 0.14 to 0.47 depending on imaging modality.\(^26,27,28\) Despite being one of the smallest reported variations, it is still larger than the patient set-up error and/or organ motion.

In prostate interobserver variability for CTV delineation seems to be smaller than in other tumour sites with $V_{\text{max}}/V_{\text{min}}$ ratio from 1.2 to 1.6, which is probably due to a better circumscribed CTV.\(^21,26\)
The largest variation is described at the apex and the base of the prostate.\(^29,40\) Valicenti et al.\(^21\) found that interobserver variability is 4 times larger for seminal vesicles delineation compared to prostate delineation.
For breast cancer the largest interobserver variability is reported for lumpectomy cavity with CI from 0.19 to 0.56, followed by CTV with CI from 0.38 to 0.87 and PTV with CI from 0.45 to 0.92.\(^11,13,41,42\)
In partial breast brachytherapy CI for lumpectomy cavity ranges from 0.48 to 0.52 and for PTV from 0.55 to 0.59, with $V_{\text{max}}/V_{\text{min}}$ ratio for all volumes 2.2 – 2.8.\(^43\) Lower CI for lumpectomy cavity compared to other target volumes could be attributed to the fact that lumpectomy cavity is the smallest target volume in postoperative breast carcinoma and CI is sensitive to volume size.
How described interobserver variability affected delivered dose to the target and/or OAR is only reported in a few papers. Steenbakkers et al.\(^44\) observed a reduction of mean dose to the rectal wall by 5.1 Gy and to the penile bulb by 11.6 Gy when reducing interobserver variability by using MRI for delineation in EBRT for prostate cancer. Allowing the same dose to OAR as in CT based delineation the dose prescribed to the target volume (prostate) could be escalated from 78 to 85 Gy. With improved target volume delineation due to the use of CT/MRI fusion in nasopharyngeal carcinoma, the mean PTV $D_{95}$ improved from 60 to 69.3 Gy, while $D_{5}$ to the brainstem and spinal cord was reduced by 19%, dose to the parotid glands and cochlea was reduced below their dose constraint.\(^45\) In lung cancer the probability of delivering at least 95% of prescribed dose to at least 95% of the target volume was reduced from 96% to 88% when using a plan designed to cover another observer’s GTV. Mean interobserver range of irradiated normal tissue volume was $12\%$, with a maximum variability.
of 66%.

Strategies to improve target volume delineation

Several strategies to reduce uncertainties in target volume delineation have been proposed by different authors and there have been a few attempts to implement those strategies to improve quality assurance in clinical trials in radiation oncology. Three major areas that could contribute to improving the accuracy of target delineation have been identified: optimisation of imaging, implementation of standardized protocols and delineation guidelines and specialized training.

Optimisation of imaging

High quality imaging with reproducible protocols is a pre-requisite for accurate target volume delineation. In the last decades, radiotherapy planning was mostly CT based, recently, new imaging techniques i.e. MRI, PET-CT, functional MRI are increasingly being used to improve visibility of the target. Potential advantages of functional imaging modalities are reduction of interobserver variability, identification of tumour extensions missed by CT and/or MRI and possibly identification of GTV subvolumes requiring higher radiation dose. Even in the absence of modern imaging modalities for treatment planning, simple measures such as the use of intravenous and/or intracavitary contrast, fiducial markers and reproducible imaging protocols can markedly increase the quality of imaging. When contouring, the use of zoom levels, simultaneous viewing in multiple planes (use of sagittal and coronal plane) and use of adequate level and window settings on the planning CT reduce interobserver variability.

In a series of 42 patients with rectal cancer, the use of PET-CT significantly reduced the size of GTV compared to CT alone, better interobserver agreement was observed (mean CI 0.79 vs. 0.82 vs. 0.93 for CT, PET-CT and PET-CT with auto-contours, respectively). Additionally, in almost one third of patients GTV based on PET-CT extended outside CT based GTV. The addition of MRI to CT did not result in significant improvement of CI. Patel et al. also compared CT and PET-CT for delineation of primary and nodal GTV (GTVp and GTVn) in rectal cancer. Similarity index for GTVp was modestly better, but statistically significant on PET CT e.g. 0.81 vs. 0.77, and notably better for GTVn e.g. 0.70 vs. 0.22. Several studies show a good correlation between PET-CT and pathology based tumour length in oesophageal cancer, but to our knowledge, there are no studies comparing interobserver variability on different imaging modalities. The benefit of PE-CT based delineation was also demonstrated for GTV in lung cancer patients, where registration of PET-CT images with the planning CT improved median interobserver percentage of concordance form 61% to 70% compared to CT alone. In RTOG 0515 trial the lung cancer GTV volumes contoured on PET CT were significantly smaller when compared to CT derived volumes and nodal GTV was altered in over 50% of patients on PET-CT. When compared to pathological findings both CT and PET-CT based contours overestimated tumour size for 46.6% and 32.5%, respectively. Both GTV volumes and maximal tumour diameters were larger on CT.

There are several publications evaluating the effect of addition of PET-CT and/or MRI on interobserver variability in delineation of the GTV or CTV for head and neck tumours. In a study by Daisne et al. GTV was contoured on CT, MRI and
PET-CT in 29 patients with head and neck tumours. Mean GTV volume was not significantly different on CT and MRI, mean GTVs on PET-CT were significantly smaller. For nine patients where surgical specimen after total laryngectomy was available, no imaging modality adequately depicted the extension of the tumour. The average GTVs for anatomic imaging were over 100% larger and for functional imaging almost 50% larger than the surgical specimen. For laryngeal and hypopharyngeal tumours mean GTV volume was 21.4 cm³ for both CT and MRI, 16.4 cm³ for PET-CT and 12.6 cm³ in surgical specimen. PET-CT was the most accurate modality in patients where comparison with the surgical specimen was available. In a similar comparison Thiagajaran et al.⁹ compared contouring GTVs in oropharyngeal carcinoma on CT + PET vs. CT + MR vs. CT + PET + MR to a reference contour and found no significant difference in the size of the GTV when contouring using any combination of two imaging modalities. Interobserver agreement between GTV CT + PET and GTV CT + MR was low, with CI = 0.62. When compared to the reference contour CI CT + PET + MR was low (0.62), but still significantly higher than CI for either CT + PET or CT + MR (0.54 and 0.55, respectively), which implicates that none of the imaging modalities should be used alone. For nodal GTV CI was > 0.75 for all tested imaging modalities compared to the reference contour, the added benefit to contrast enhanced CT alone was small. Anderson et al.⁵⁸ also compared CT, PET-CT and MRI for definition of GTV in head and neck tumours. Interobserver variability was present for all imaging modalities, with CT being least consistent. PET-CT derived target volumes were the smallest in size, interobserver agreement was the highest with CI = 0.46, compared to CI = 0.36 and 0.35 for MRI and CT, respectively. In nasopharyngeal carcinoma the use of CT and co-registered MRI decreased local SD from 4.4, to 3.3 mm and from 5.9 to 4.9 mm for CTV and elective CTV, respectively, and resulted in a higher agreement between observers.²⁴ Two published studies observed no significant difference between observers across imaging modalities when comparing CT to PET-CT and CT to MRI for GTV delineation in head and neck tumours.⁵⁶,⁵⁷ Giezen et al.⁵¹,⁵² compared CT and MRI for delineation of CTV and lumpectomy cavity (LC) after breast-conserving surgery and found that both imaging modalities provided similar visibility of LC, CI was lower for MRI than for CT, but the difference was not significant. These results have to be interpreted with caution, as the participating radiologists had no experience in LC contouring and the radiation oncologists were not familiar with breast MRI, which gives the results limited value. In postoperative brain gliomas radiotherapy the use of registered CT and postoperative MR images reduced interobserver variability compared to contouring on CT with the aid of preoperative MRI (CI 0.47 vs. 0.14, respectively). However, in delineation of inoperable supratentorial brain tumours the addition of MRI did not reduce interobserver variability with Vmax remaining up to 2.7 times larger than Vmax.³⁸ For prostate cancer all studies demonstrate up to almost 75% larger volumes on CT compared to MRI, but while some found better interobserver agreement on MRI others found less interobserver variability on CT, demonstrating that current delineation guidelines might not be applicable to MRI planning.⁴⁰,⁵⁹,⁶⁰

Implementation of delineation protocols and guidelines

Delineation guidelines have been published on a national or international level for several tumour sites both in EBRT and BT.⁵¹-⁵⁷ Different reports show that the use of site specific anatomical atlases, consensus delineation guidelines and standardized contouring protocols diminish variability between observers in various tumour sites.³²,⁶⁸-⁷¹ In rectal carcinoma, the implementation of site specific consensus atlas significantly reduced interobserver variability in a pilot study⁶⁹, which was later confirmed in a larger study, in which Nijkamp et al.³² demonstrated that the use of a digital delineation atlas twice or more during target volume contouring significantly improves CI. The addition of delineation guidelines significantly reduced interobserver variation in caudal CTV border (from 1.8 to 1.2 cm) and the size of average CTV volume by 25% (620 vs. 460 cc). In lung cancer, re-contouring of the GTV with the use of a protocol, aimed at minimizing variation, reduced the degree of interobserver variation from 20% to 13%. In the second contouring session the differences between observers were not statistically significant.⁷² Comparison of contouring seroma cavity in partial breast radiotherapy with and without guidelines showed that radiation oncologists (ROs) contouring without guidelines contoured significantly larger CTVs and PTVs in more than 50% of patients.⁶⁸ When all participating ROs were provided with guidelines, the differences in sizes of the target volumes were no longer significant. In breast brachytherapy conformity indices increased significantly with the use of guidelines.
both for lumpectomy cavity contours and PTV. The increase was 14% and 11% for the cavity and 28% and 17% for PTV on preimplant and post-implant CT images, respectively. Even for site-specialized ROs, a reduction in interobserver variability was noticed in CTV delineation for postprostatectomy radiotherapy when adhering to the RADICALS trial delineation protocol. Mean $V_{\text{max}}/V_{\text{min}}$ for all cases was reduced from 3.7 at first delineation to 2.0 at the second delineation.

<table>
<thead>
<tr>
<th>Tumour site</th>
<th>Target volume</th>
<th>No of pts</th>
<th>No of obs</th>
<th>Imaging modality</th>
<th>Results</th>
<th>Author (publication date)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectum</td>
<td>GTV, CTV</td>
<td>2</td>
<td>10</td>
<td>CT, PETCT</td>
<td>CI(GTV) = 0.26-0.33 CI(CTV) = 0.29-0.35</td>
<td>Krengli et al., 2010</td>
</tr>
<tr>
<td></td>
<td>GTV</td>
<td>52</td>
<td>5</td>
<td>CT, PETCT, MRI</td>
<td>CI = 0.79-0.93</td>
<td>Buijsen et al., 2012</td>
</tr>
<tr>
<td></td>
<td>CTV</td>
<td>8</td>
<td>10</td>
<td>CT</td>
<td>CI = 0.63-0.66</td>
<td>Nijkamp et al., 2012</td>
</tr>
<tr>
<td></td>
<td>GTV</td>
<td>6</td>
<td>4</td>
<td>CT, PETCT</td>
<td>SI(GTV-P) = 0.77-0.81 SI(GTV-N) = 0.22-0.70</td>
<td>Patel et al., 2007</td>
</tr>
<tr>
<td>Stomach</td>
<td>CTV, PTV</td>
<td>1</td>
<td>10</td>
<td>CT</td>
<td>Vmax/Vmin(CTV) = 3.4 Vmax/Vmin(PTV) = 2.6</td>
<td>Jansen et al., 2010</td>
</tr>
<tr>
<td>Gastroesophage</td>
<td>GTV</td>
<td>1</td>
<td>50</td>
<td>CT</td>
<td>JCl = 0.69</td>
<td>Gwynne et al., 2012</td>
</tr>
<tr>
<td></td>
<td>GTV, CTV, PTV</td>
<td>1</td>
<td>48</td>
<td>CT</td>
<td>Vmax/Vmin(PTV) = 5.25-6.03</td>
<td>Tai et al., 1998</td>
</tr>
<tr>
<td>Cervix EBRT</td>
<td>CTV</td>
<td>3</td>
<td>7</td>
<td>CT</td>
<td>CI = 0.11-0.57 Vmax/Vmin = 3.6-4.9</td>
<td>Weiss et al., 2003</td>
</tr>
<tr>
<td>IGABT</td>
<td>GTV, HRTC, IRCTV</td>
<td>6</td>
<td>10</td>
<td>MRI</td>
<td>CI(GTV) = 0.6-0.8 CI(HR&amp;RCVT) = 0.6-0.7</td>
<td>Petrč et al., 2012, 2013</td>
</tr>
<tr>
<td>Head and neck</td>
<td>GTV, CTV, PTV</td>
<td>1</td>
<td>20</td>
<td>CT</td>
<td>Vmax/Vmin(CTV) = 18.3</td>
<td>Hong et al., 2012</td>
</tr>
<tr>
<td></td>
<td>GTV</td>
<td>41</td>
<td>3</td>
<td>CT, PETCT, MRI</td>
<td>CI(GTV-P) = 0.54-0.62 CI(GTV-N) = 0.75</td>
<td>Thiagajaran et al., 2012</td>
</tr>
<tr>
<td></td>
<td>CTV, CTVE</td>
<td>10</td>
<td>10</td>
<td>CT, MRI</td>
<td>localSD(CTV) = 3.3-4.4mm localSD(CTVe) = 4.9-5.9mm</td>
<td>Rasch et al., 2012</td>
</tr>
<tr>
<td>Lung</td>
<td>GTV</td>
<td>12</td>
<td>8</td>
<td>CT, CBCT</td>
<td>CI = 0.27-0.39 Cigen = 0.56-0.70</td>
<td>Altorjai et al., 2012</td>
</tr>
<tr>
<td></td>
<td>GTV</td>
<td>8</td>
<td>5</td>
<td>CT</td>
<td>Vmax/Vmin&gt;7</td>
<td>Van De Steene et al., 2002</td>
</tr>
<tr>
<td></td>
<td>GTV</td>
<td>10</td>
<td>17</td>
<td>CT</td>
<td>Vmax/Vmin = 5.2 CI = 0.04-0.48</td>
<td>Graud et al., 2002</td>
</tr>
<tr>
<td></td>
<td>GTV</td>
<td>22</td>
<td>11</td>
<td>CT, PETCT</td>
<td>meanCI = 0.17[CT], 0.09[PETCT] localSD = 1cm[CT], 0.4cm[PETCT]</td>
<td>Steenbakers et al., 2006</td>
</tr>
<tr>
<td></td>
<td>GTV</td>
<td>19</td>
<td>2</td>
<td>CT, PETCT</td>
<td>medianCI(CT) = 0.6, medianCI(PTV) = 0.70</td>
<td>Fox et al., 2005</td>
</tr>
<tr>
<td>Brain</td>
<td>CTV</td>
<td>7</td>
<td>5</td>
<td>CT + MRI</td>
<td>CI = 0.14-0.47</td>
<td>Cattaneo et al., 2005</td>
</tr>
<tr>
<td></td>
<td>GTV</td>
<td>5</td>
<td>9</td>
<td>CT, MRI</td>
<td>Vmax/Vmin(CT) = 1.7-2.8 Vmax/Vmin(MR) = 1.5-2.7</td>
<td>Weltens et al., 2001</td>
</tr>
<tr>
<td>Prostate</td>
<td>Prostate, seminal vesicles (SV)</td>
<td>10</td>
<td>7</td>
<td>CT</td>
<td>Vmax/Vmin(P) = 1.18-1.63 Vmax/Vmin(SV) = 2.02-6.43</td>
<td>Valicenti et al., 1999</td>
</tr>
<tr>
<td>Prostate</td>
<td>Prostate</td>
<td>3</td>
<td>2</td>
<td>CT</td>
<td>Vmax/Vmin = 1.39-1.65</td>
<td>Seddon et al., 2000</td>
</tr>
<tr>
<td>Prostate</td>
<td>Prostate</td>
<td>5</td>
<td>5</td>
<td>CT, MRI</td>
<td>MeanCI(MR) = 0.83 MeanCI(C) = 0.69</td>
<td>Segedin et al., 2011</td>
</tr>
<tr>
<td>Breast</td>
<td>Lumpectomy cavity (LC), CTV</td>
<td>15</td>
<td>3</td>
<td>CT, MRI</td>
<td>CI(LC) = 0.32[MR], 0.52(C) CI(CTV) = 0.77[MR], 0.29(C)</td>
<td>Giezen et al., 2011, 2012</td>
</tr>
<tr>
<td></td>
<td>Lumpectomy cavity</td>
<td>30</td>
<td>5</td>
<td>CT</td>
<td>MeanCI = 0.36</td>
<td>Boersma et al., 2012</td>
</tr>
<tr>
<td></td>
<td>Lumpectomy cavity, PTV</td>
<td>8</td>
<td>13</td>
<td>CT</td>
<td>CI(LC) = 0.19-0.77 CI(CTV) = 0.38-0.80 CI(PTV) = 0.45-0.61</td>
<td>VanMourik et al., 2010</td>
</tr>
<tr>
<td></td>
<td>Lumpectomy cavity, PTV</td>
<td>9</td>
<td>5</td>
<td>CT</td>
<td>CI(LC) = 0.48-0.52 CI(PTV) = 0.55-0.59 Vmax/Vmin = 2.2-2.8</td>
<td>Majo et al., 2015</td>
</tr>
<tr>
<td></td>
<td>Lumpectomy cavity, CTV</td>
<td>18</td>
<td>5</td>
<td>CT</td>
<td>MeanCI(LC) = 0.56 MeanCI(CTV) = 0.87</td>
<td>Struikmans et al., 2005</td>
</tr>
</tbody>
</table>

Cl = conformity/concordance index; CTV = clinical target volume; GTV = gross target volume; local SD = local standard deviation; max = maximum; min = minimum; obs = observers; pts = patients; PTV = planning target volume; SI = similarity index; V = volume.

$\frac{V_{\text{max}}}{V_{\text{min}}}$ for all cases was reduced from 3.7 at first delineation to 2.0 at the second delineation.
Training

A survey of radiotherapy planning and delivery undertaken in the UK in 2007 showed a lack of formal education in target volume and OAR delineation in different staff groups. Only 4% of NHS radiotherapy departments offered structured training on image interpretation, while 6% offered informal sessions with radiologists. 90% of participating ROs stated they wanted formal training in interpretation of cross sectional imaging and almost 85% were interested in online training modules. More than half of junior ROs considered their training in cross sectional imaging to be inadequate.

Some publications evaluated the effect of clinical experience on interobserver variability, the results, however, were ambiguous. While Hurkmans et al. reported that more experienced ROs delineate smaller volumes than unexperienced in breast carcinomas, Giraud et al. found experienced ROs to delineate larger volumes than their younger colleagues in lung carcinoma. In brain tumours, Leunens et al. found no significant difference between experienced and unexperienced ROs.

Only a few publications have addressed the subject of training, some in the course of pre-accrual quality assurance delineation exercises (dummy run). In dummy run for a randomised multicentre PET-plan clinical trial in lung cancer, they found considerable differences despite providing detailed contouring guidelines. After a teaching session at a study group meeting, they observed an improvement in overall interobserver agreement, demonstrated by reduction of target volumes and an increase in kappa (κ) indices for GTV and two CTVs (0.63 vs. 0.71, 0.60 vs. 0.65 and 0.59 vs. 0.63, respectively). Similarly, Khoo et al. reported reduced encompassing to intersecting volume ratio (VR) at re-contouring the prostate after education sessions focusing on MRI prostate anatomy with CT correlation. Mean VR was reduced by 15% for CT (from 2.74 to 2.33) and 40% for MRI (from 2.38 to 1.41). Dewas et al., however, found no significant difference for delineation of the target volumes in lung cancer before and after training. The residents κ-indices were lower compared to senior ROs both before and after the training, V20 for lung was higher in the residents group. The authors speculated there was no improvement because initial delineations by the residents were good. However, they offered no hands-on training for the residents and most reports showing improvement included hands-on training in their educational sessions. During training, special attention needs to be paid to predilection areas for larger interobserver variability, identified in available literature.

Conclusions

The main goal of improving accuracy in radiotherapy treatment planning and delivery is better local control with less morbidity, resulting in better quality of life. Our review shows that interobserver variability in target volume contouring represents the largest uncertainty in the process for most tumour sites, potentially resulting in geographic miss in dose delivery, which could hamper local control for individual patients. Studies on use of modality imaging and image co-registration show promising results, however, for most tumour sites the optimal combination of imaging modalities still needs to be determined. Strict introduction and use of imaging and delineation protocols and guidelines reduces interobserver variability, therefore it is advisable in every day practice and mandatory in the frame of clinical studies. Especially in multicentric studies, efforts to unify target volume delineation in different institutions in a dummy run should be maximized as interobserver variability influences reliability of dose reporting, comparison of treatment outcomes and interpretation of study results, hence diminishing the value of a study. To assure adherence to study protocols and delineation guidelines, a central reviewing board for contour correction is useful. Continuing medical education of ROs cannot be overemphasized, intensive formal training on interpretation of sectional imaging should be included in the program for radiation oncology residents. In the fields, where the other conditions are fulfilled (recommendations on imaging for treatment planning, delineation guidelines), a study systematically assessing the effect of training on interobserver variability is warranted.

Acknowledgments

Authors would like to thank Robert Hudej for preparing the figures for this manuscript.

References


6. Njeh CF, Dong L, Orton CG. IGRT has limited clinical value due to lack of accuracy. Radiother Oncol 2010; 94: 286-91.


