Interobserver variability in clinical target volume delineation for primary mediastinal B-cell lymphoma

Cristina Piva MD a, Domenico Genovesi MD b, Andrea Riccardo Filippi MD a,⁎, Mario Balducci MD c, Salvina Barra MD d, Michela Buglione MD e, Mario Busetto MD f, Patrizia Ciammella MD g, Paola Franzone MD h, Vitaliana De Sanctis MD i, Gabriele Simontacchi MD j, Marco Fusella MSc k, Umberto Ricardi MD a on behalf of Radiation Oncology Committee of the Italian Lymphoma Foundation (FIL)

a Department of Oncology, University of Torino, Turin, Italy
b Department of Radiation Oncology, University of Chieti, SS Annunziata Hospital, Chieti, Italy
c Radiotherapy Department, Radiation Oncology, Catholic University, Rome, Italy
d Radiation Oncology, Istituto Nazionale per la Ricerca sul Cancro, Genoa, Italy
e Radiation Oncology, University of Brescia, Brescia, Italy
f Radiotherapy Unit, Ospedale dell’Angelo, Mestre, Italy
g Department of Oncology and Advanced Technology, Radiation Therapy Unit, Arcispedale S. Maria Nuova Hospital-Istituto di Ricovero e Cura a Carattere Scientifico, Reggio Emilia, Italy
h Radiation Oncology, Azienda Ospedaliera SS Antonio, Biagio e Cesare Arrigo, Alessandria, Italy
i Radiation Oncology, Azienda Ospedaliera S. Andrea, La Sapienza University, Rome, Italy
j Radiotherapy, University of Florence, Florence, Italy
k Medical Physics, University of Torino, Turin, Italy

Received 13 February 2015; revised 7 April 2015; accepted 10 April 2015

Abstract

Purpose: The purpose of this study was to evaluate interobserver variability among radiation oncologists with experience in the field of lymphoma radiation therapy in the delineation of clinical target volume (CTV) in a challenging case of primary mediastinal B-cell lymphoma.

Methods and materials: Ten experienced radiation oncologists were invited to a 1-day contouring session. The case of a 56-year-old man with primary mediastinal B-cell lymphoma with complete metabolic response after chemotherapy was chosen as the sample for the study. A brief presentation of his clinical history was given, together with guidelines for contouring. The 10 CTVs obtained were then compared in terms of variation in total volume and in craniocaudal, laterolateral, and anteroposterior diameters. The CTV with the best Dice similarity coefficient

Conflicts of interest: None.

⁎ Corresponding author. University of Torino, Department of Oncology, Via Genova 3, 10126 Turin, Italy.
E-mail address: andreariccardo.filippi@unito.it (A.R. Filippi).

Note—Earn CME credit by taking a brief online assessment at https://www.astro.org/JournalCME.
(DSC) between the union of all 10 CTVs and the individual CTV was considered the reference CTV, and the DSC and the Hausdorff distance (HD) for each volume compared with the reference CTV were then calculated.

Results: A significant variability was found in total volume (mean, 498.3 cm$^3$; range, 181.8-1003 cm$^3$) and craniocaudal (median, 144.7 mm; range, 80.6-159 mm), laterolateral (median, 133.5 mm; range, 83.7-149.5 mm), and anteroposterior diameters (median, 136.2 mm; range, 84-150.5 mm). Analysis of the DSC and the HD showed a mean DSC of 0.53 (range, 0.31-0.74) and a mean HD of 6.4 cm (range, 1.8-14.8 cm).

Conclusions: Results of this study strongly indicate the need to develop and share appropriate contouring guidelines among experts and suggest the promotion of specific educational activities to improve radiation therapy quality in both clinical trials and routine clinical practice.

© 2015 American Society for Radiation Oncology. Published by Elsevier Inc. All rights reserved.

Introduction

Primary mediastinal B-cell lymphoma (PMBCL) is a distinct and uncommon variant of diffuse large B-cell lymphoma (DLBCL) arising from thymic B cells. The therapeutic “gold standard” is currently represented by a combination of chemotherapy (CHOP [cyclophosphamide, doxorubicin, vincristine, and prednisone]; CHOP-like; EPOCH [etoposide, prednisone, vincristine, cyclophosphamide, doxorubicin]; or M/VACOP-B [methotrexate/etoposide, doxorubicin, cyclophosphamide, vincristine, prednisone, and bleomycin]) and rituximab followed by radiation therapy (RT). The need for consolidation RT has been questioned for patients with complete response after rituximab chemotherapy (R-CT), and the prognostic role of postchemoimmunotherapy ($^{18}$F) fluorodeoxyglucose positron emission tomography ($^{18}$F]FDG-PET) is under evaluation.

An ongoing randomized trial led by the collaborative International Extranodal Lymphoma Study Group, the IELSG 37 study, will assess the role of consolidation RT in patients with PMBCL with complete metabolic response (CMR) after conventional chemoimmunotherapy. In the IELSG 37 study, for patients with a positive $^{18}$F]FDG-PET finding after R-CT, the decision about further treatment is left to the treating physicians, whereas all patients with a negative $^{18}$F]FDG-PET result after R-CT are randomized to either receive consolidation mediastinal RT (30 Gy) or undergo observation. Of note, the IELSG 37 study has a quality assurance program for RT with a centralized review of contouring and planning, to minimize potential biases secondary to radiation planning and delivery.

The mediastinum is a challenging structure for radiation oncologists. A few studies have shown that imaging interpretation may be extremely different when the mediastinum is involved in lung cancer and in Hodgkin lymphoma (HL). Moreover, interobserver variability in the delineation of gross tumor volume, clinical target volume (CTV), and planning target volume represents a significant problem for RT planning. For this reason, the IELSG 37 study emphasizes the importance of defining all initially involved disease sites with information from prechemotherapy imaging and delineating the CTV as the initial volume of the mediastinal mass at presentation, taking into account response to chemotherapy and displacement of normal structures.

The aim of this study, coordinated by the Radiotherapy Working Party of the Italian Lymphoma Foundation (FIL), was to assess interobserver variability among expert radiation oncologists in the field of lymphoma RT in CTV definition of PMBCL, to provide data that could be used as the starting point for the development of specific contouring guidelines and related educational activities.

Methods and materials

The case of a 56-year-old man affected by PMBCL with disease involving the anterior mediastinum, with extension to the middle mediastinum, anterior chest wall, right costophrenic angle, and bilateral peribronchial lymph nodes, was presented. The patient received 6 cycles of CHOP plus rituximab chemotherapy every 14 days (R-CHOP-14), achieving a CMR according to the Cheson criteria. Study participants were members of a national working party on RT for hematologic malignancies, and they all worked in hospitals or academic centers with a high number of lymphoma patients each year. They had complete clinical data on patient history, computed tomography (CT) and CT-PET scans at the time of diagnosis (Fig 1) and after chemotherapy (Fig 2), and treatment planning scans. The treatment planning CT was obtained with the subject in a supine position with the use of a specific immobilization device on a multislice CT scanner with 0.5-cm slice thickness and without intravenous contrast medium. All participants were invited to delineate the CTV according to specific guidelines obtained from the IELSG study protocol for IELSG 37 (ClinicalTrials.gov unique identifier NCT01599559). Participants were informed on clinical history and on IELSG 37 contouring guidelines the same day as the meeting and were assigned to 2 separate workstations equipped with deformable imaging registration software (VelocityAI, v2.7; Velocity Medical Solutions, Atlanta, Atlanta,
Through the use of the Velocity software, every observer had the opportunity to look at both prechemotherapy and postchemotherapy CT scans (with intravenous contrast medium) and prechemotherapy and postchemotherapy CT-PET images. The different sets of images were then coregistered for delineation with importation of RT structures. Every participant had 2 hours to complete the CTV contouring.

Figure 1 Graphic illustration on axial and coronal planes of disease presentation at the time of diagnosis on computed tomography (A-E) and CT-PET (F-I, L) scans.

Figure 2 Graphic illustration on axial (A-D) and coronal (E) planes of disease presentation after chemotherapy on computed tomography scan.
Data Analysis

Each CTV was measured in cubic centimeters; the maximum craniocaudal, laterolateral, and anteroposterior diameters of each CTV were calculated in millimeters. Minimum and maximum observation mean value, standard deviation, and median value were calculated for each variable (volume and CTV diameters).

We considered as the CTV reference (CTV\textsubscript{ref}) the one with the best Dice similarity coefficient (DSC) between a

Figure 3  Graphic representation of different axial and coronal planes of interobserver variation among 10 observers.

Figure 4  Variability of volume for clinical target volumes delineated by each observer.
volume that resulted from the union of all CTVs and the individual CTV. The DSC for each volume compared with the CTV_ref was used to describe interobserver concordance, and the mean DSC was calculated. A DSC of 0 indicates no agreement between the observers, and a DSC of 1 represents 100% agreement. Furthermore, to evaluate interobserver concordance, we also used the Hausdorff Distance (HD), which is a mathematical construct used to measure the “closeness” of 2 sets of points. It determines the maximum of all the distances from points on 1 structure to the closest point on the other structure. Low values of HD indicate no outlier points on the structures being compared.

Results

A graphic representation of interobserver variability in CTV definition is shown on a sample CT axial and coronal plane in Fig 3. Variability in CTV definition was remarkable, with CTV volumes ranging from 181.8 to 1003 cm³, as illustrated in Fig 4. The mean CTV was 498.3 cm³, with a standard deviation of 285.5 cm³. These variations were confirmed when we measured craniocaudal (median, 144.7 mm; range, 80.6-159 mm), anteroposterior (median, 136.2 mm; range, 84-150.5 mm), and laterolateral (median, 133.5 mm; range, 83.7-149.5 mm) diameters, as shown in Table 1. Figure 5 shows the variability of craniocaudal, anteroposterior, and laterolateral diameters drawn by each observer on the CT slices. Table 2 provides DSC and HD values obtained by the comparison between CTV_ref and each CTV. Mean DSC and HD values were 0.53 (range, 0.31-0.74) and 6.4 cm (range, 1.8-14.8 cm), respectively.

Discussion

The present study was designed with the aim of evaluating the interobserver variability among a group of expert radiation oncologists, who specialized in lymphoma RT, in a challenging case of PMBCL. The study was part of the activities of a network between several Italian radiation oncology departments coordinated by the Radiation Oncology Committee of the FIL. The main goal of the working party is to reduce individual variations in radiation indications, target delineation, dose prescription, and treatment technique for lymphoma patients. These activities are specifically dedicated to patients enrolled in FIL clinical trials that include RT or in trials run by other scientific groups endorsed by FIL, such as the IELSG37, but it is hoped that this type of work will be useful for increasing the global quality of radiation treatment for hematologic malignancies all over the world.

Table 1  Volumetric and dimensional results

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Volume (cm³)</th>
<th>Craniocaudal diameter (mm)</th>
<th>Anteroposterior diameter (mm)</th>
<th>Laterolateral diameter (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Mean</td>
<td>498.3</td>
<td>132.7</td>
<td>132.6</td>
<td>126.2</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>285.5</td>
<td>24.9</td>
<td>18.9</td>
<td>21.6</td>
</tr>
<tr>
<td>Minimum</td>
<td>181.8</td>
<td>80.6</td>
<td>84</td>
<td>83.7</td>
</tr>
<tr>
<td>Maximum</td>
<td>1003</td>
<td>159</td>
<td>150.5</td>
<td>149.5</td>
</tr>
<tr>
<td>Median</td>
<td>430.4</td>
<td>144.7</td>
<td>136.2</td>
<td>133.5</td>
</tr>
</tbody>
</table>

Figure 5  Box-and-whisker plots for maximum anteroposterior (AP), craniocaudal (CC), and laterolateral (LL) diameters.
country. Precise and accurate target delineation is crucial in lymphoma RT (and in light of the dramatic changes in volumes and treatment techniques that have occurred over time) but is highly dependent on imaging interpretation and may generate errors.\(^\text{10,14}\) Compared with other tumors, for lymphoma mediastinal presentations, the anatomic changes, tumor shrinkage after chemotherapy, different positioning during diagnostic and treatment planning procedures, and respiratory organ motion can all seriously complicate a precise target definition.\(^\text{15}\) Interobserver variability in CTV delineation may also be caused by a diverging evaluation of the exact localization of lymph node sites. Several studies have shown that radiologists assess CT images differently to determine the mediastinal lymph node status of patients affected with lung cancer.\(^\text{7,16,17}\) Fletcher et al\(^\text{8}\) found that expert mediastinal lymph node status of patients affected with CT images of HL. Several authors have suggested that the registration of PET-CT scans with RT planning CTs may reduce interobserver variability with regard to the accuracy of CTV definition in head and neck cancer,\(^\text{18}\) adult and pediatric HL,\(^\text{19,20}\) lung cancer,\(^\text{21}\) and thoracic non-HL.\(^\text{22}\) Some studies also evaluated the interobserver variability in target volume delineation in supradiaphragmatic adult HL and pediatric HL using parameters such as total volume, diameters, and conformity index.\(^\text{12,23}\) Genovesi et al\(^\text{12}\) proposed a multi-institutional contouring dummy run of 2 cases of early-stage supradiaphragmatic HL that were contoured by 18 and 15 observers, respectively. Both cases presented significant variability in total volume (range, 74.1–1157.1 cm\(^3\) and 341.8–1662 cm\(^3\), respectively), and these variations were confirmed when craniocaudal, anteroposterior, and laterolateral diameters were measured.\(^\text{12}\) Lutgendorf-Caucig et al\(^\text{23}\) performed a multicenter evaluation of different target volume delineation concepts (involved node vs involved node level) in 2 cases of pediatric HL. Seven radiation oncologists contoured neck and mediastinal CTVs, showing better agreement in both cases for the involved node level concept than the involved node concept. Nevertheless, they obtained low conformity index values (0.28 vs 0.39, 0.14 vs 0.24, and 0.18 vs 0.33 for mediastinum, right neck, and left neck, respectively).\(^\text{23}\)

Recently, Hoppe et al\(^\text{24}\) conducted a survey regarding the correct interpretation of the International Lymphoma Radiation Oncology Group involved-site RT guidelines, showing great variability even among expert radiation oncologists. To the best of our knowledge, this is the first study investigating interobserver variability in PMBCL.

For study purposes, the clinical case was selected as presenting the typical challenging characteristics of a PMBCL, with a bulky mediastinal mass and the involvement of such nodal sites as the anterior chest wall, the right costophrenic angle, and bilateral peribronchial lymph nodes. Despite available guidelines provided contextually at the meeting, great variability was recorded among observers from different institutions. The main variations in diameters were observed in the craniocaudal direction, with a minimum value of 80.6 mm and a maximum value of 159 mm. Mean CTV and standard deviations in terms of volume were 498.3 and 285.5 cm\(^3\), respectively. The great discrepancy in the CTVs was confirmed by the DSC values, a parameter used in some similar studies to evaluate interobserver variability in pediatric HL\(^\text{23}\) and in pancreatic cancer.\(^\text{25}\) In our study, the DSC ranged between 0.31 and 0.74, with a mean value equal to 0.53, which does not represent much agreement between observers. Furthermore, the HD, with a mean value of 6.4 cm, confirmed the low agreement between observers.

Notably, despite the fact that coregistration of the PET-CT scan with the RT planning CT should facilitate CTV contouring, as evidenced by several studies,\(^\text{18-22}\) not all observers had the same level of experience with software that allowed image fusions, and in particular with the software used here (VelocityAI, v2.7). This may explain in part the high variability in target delineation. Another critical issue that may have had an impact on the variability in CTV definition is the different clinical approach used among observers, despite the guidelines. In fact, as already mentioned, the interpretation of contouring guidelines may vary widely among experts.\(^\text{24,26}\) In the present study, we noted that some observers delineated all of the initially involved sites of disease, taking into account the response to chemotherapy and the displacement of normal structures according to an involved-site RT approach.\(^\text{27}\) In contrast, other observers defined a smaller volume according to the concept of involved nodal RT\(^\text{28}\) or the residual mass only.

Several studies have demonstrated that a reduced number of RT protocol violations might have an impact on survival. A retrospective quality control study by the German Hodgkin Study Group showed that deviations in the definition of radiation treatment fields and prescription doses from prospective treatment guidelines were unfavorable prognostic factors.\(^\text{29}\) For these reasons, the IELSG 37 trial requires that RT data concerning treatment volumes and dosimetric parameters be uploaded within 1 month after completion of RT to permit a retrospective quality assurance program.\(^\text{6}\)
In conclusion, this study showed that despite the observers’ experience, the availability of a correct prechemotherapy and postchemotherapy imaging study, and specific contouring guidelines, variability in CTV delineation among observers was extremely high. The Radiation Oncology Committee of the FIL is currently working on the implementation of national guidelines for target delineation, dose prescription, and planning in various nodal and extranodal lymphomas to reduce this variability and increase the global quality of radiation treatment. An initial dedicated contouring workshop on lymphomas was proposed during the national meeting in November 2014 (the national meeting was the AIRO (Associazione Italiana Radioterapia Oncologica) meeting), and others will be planned over the next several months. In designing clinical trials, investigators should be aware of the wide variability in target definitions for RT, and a prospective/retrospective quality assurance program should be implemented.

References