

Interobserver delineation uncertainty in Involved Node Radiation Therapy (INRT) for early-stage Hodgkin lymphoma: on behalf of the Radiotherapy Committee of the EORTC lymphoma group

Marianne C Aznar<sup>1</sup>, Theodore Girinsky<sup>2</sup>, Anne Kiil Berthelsen<sup>1</sup>, Berthe Aleman<sup>3</sup>, Max Beijert<sup>4</sup>, Martin Hutchings<sup>5</sup>, Yolande Lievens<sup>6</sup>, Paul Meijnders<sup>7</sup>, Peter Meidahl Petersen<sup>1,5</sup>, Deborah Schut<sup>1</sup>, Maja V Maraldo<sup>1</sup>, Richard van der Maazen<sup>8</sup>, Lena Specht<sup>1,5</sup>

- 1) Department of Oncology, section of Radiotherapy, Rigshospitalet, University of Copenhagen, Denmark
- 2) Service de Radiothérapie Oncologique, Institut Gustave Roussy, France
- 3) Department of Radiotherapy, the Netherlands Cancer Institute, Amsterdam, The Netherlands
- 4) Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.
- 5) Department of Haematology, Rigshospitalet, University of Copenhagen, Denmark
- 6) Department of Radiation Oncology, Ghent University Hospital, Ghent, Belgium
- 7) Department of Radiation Oncology GZA, Iridium Cancer Network, University of Antwerp , Antwerp, Belgium
- 8) Department of Radiotherapy, Radboud University Medical Center, Nijmegen, The Netherlands

Number of pages: 13

Number of tables: 2

Number of figures: 4

Running title: Delineation uncertainty in Hodgkin Lymphoma

Keywords: delineation uncertainty, Hodgkin Lymphoma, involve-node radiotherapy

1

2 Abstract

### 3 **Background and Purpose**

4 In early stage classical Hodgkin lymphoma (HL) the target volume nowadays consists of the volume  
5 of the originally involved nodes. Delineation of this volume on a post-chemotherapy CT-scan is  
6 challenging. We report on the interobserver variability in target volume definition and its impact  
7 on resulting treatment plans.

### 8 **Materials and Methods**

9 Two representative cases were selected (1: male, stage IB, localization: left axilla; 2: female, stage  
10 IIB, localizations: mediastinum and bilateral neck). Eight experienced observers individually  
11 defined the CTV using involved node radiotherapy (INRT) as defined by the EORTC-GELA guidelines  
12 for the H10 trial. A consensus contour was generated and the standard deviation computed. We  
13 investigated the overlap between observer and consensus contour (Sørensen-Dice coefficient  
14 (DSC)) and the magnitude of gross deviations between the surfaces of the observer and consensus  
15 contour (Hausdorff distance). 3D-conformal (3D-CRT) and intensity modulated radiotherapy  
16 (IMRT) plans were calculated for each contour in order to investigate the impact of interobserver  
17 variability on each treatment modality. Similar target coverage was enforced for all plans.

### 18 **Results**

19 The median CTV was 120 cm<sup>3</sup> (IQR: 95-173 cm<sup>3</sup>) for case 1, and 255 cm<sup>3</sup> (IQR: 183-293 cm<sup>3</sup>) for  
20 case 2. DSC values were generally high (>0.7), and Hausdorff distances were about 30 mm. The  
21 standard deviations between all observer contours, providing an estimate of the systematic error  
22 associated with delineation uncertainty, ranged from 1.9 to 3.8 mm (median: 3.2 mm). Variations  
23 in mean dose resulting from different observer contours were small and were not higher in IMRT  
24 plans than in 3D-CRT plans.

### 25 **Conclusions**

26 We observed considerable differences in target volume delineation, but the systematic  
27 delineation uncertainty of around 3 mm is comparable to that reported in other tumor sites.

28 This report is a first step towards calculating an evidence-based PTV margin for INRT in HL.

29

## 1 Introduction

2 Radiotherapy (RT) for Hodgkin lymphoma (HL) has changed dramatically during the past decades.  
3 When RT was the primary treatment modality, very extensive treatment fields were used to  
4 encompass not only the macroscopic lymphoma but also possible microscopic disease. Total or  
5 subtotal nodal irradiation encompassing all the major lymph node areas was used routinely in  
6 early stage disease. With the advent of effective chemotherapy, it became clear that these large  
7 prophylactic treatment fields were no longer needed[1] and Involved field RT (IFRT), including only  
8 regions with involved lymph nodes, became the standard [2]. Soon after, studies of patients  
9 treated with chemotherapy alone showed that recurrences occurred most often at the site of  
10 initial macroscopic lymphoma involvement[3]. Using FDG-PET to identify this initial involvement,  
11 and modern 3-dimensional (3D) conformal treatment planning to target it, it became possible to  
12 reduce the treatment volume even further. The EORTC (European Organisation for the Research  
13 and Treatment of Cancer) Lymphoma Group pioneered this limited RT for early stage HL, called  
14 involved node radiotherapy (INRT)[4,5]. With INRT, the clinical target volume (CTV) includes only  
15 the volume of initially involved lymph nodes, as identified on PET/CT before chemotherapy is  
16 administered, without compromising the effectiveness of the treatment[6–8].

17 Conformal radiotherapy makes precise target definition essential in all treatment sites. In HL, and  
18 in other types of lymphomas as well, these issues are particularly challenging because of the highly  
19 variable anatomical disease localizations among patients and because of the inherent difficulty of  
20 defining the pre-chemotherapy lymphoma volume on a post-chemotherapy planning CT-scan. In  
21 addition, the CTV also includes volumes looking suspicious on CT but not PET positive. Genovesi et  
22 al[9] reported variations in CTV volumes of up to 1000 cm<sup>3</sup> among observers contouring IFRT in  
23 supra-diaphragmatic Hodgkin Lymphoma without information from FDG-PET scans and in the  
24 absence of contouring guidelines. Piva et al[10] reported similar results in a case of primary  
25 mediastinal B-cell lymphoma, despite using deformable image registration to fuse pre- and post-  
26 chemotherapy images. Though these reports have raised awareness about the challenges of  
27 delineation in Hodgkin Lymphoma, they do not reflect a “best case scenario” context in which  
28 INRT can be applied, namely having the pre-chemotherapy PET/CT scan performed in treatment

1 position in order to minimize the geometric uncertainties related to image fusion with the post-  
2 chemotherapy planning CT-scan.

3 In this study, we report on the interobserver variability in target volume delineation in close to  
4 optimal pre- and post- chemotherapy imaging conditions as defined by the published INRT  
5 guidelines[4] and using the expertise from the Radiotherapy Committee of The EORTC Lymphoma  
6 Group. We also investigate whether the dosimetric impact of this interobserver variation is greater  
7 for more conformal techniques (such as intensity modulated radiotherapy (IMRT) than for 3D-  
8 conformal (3D-CRT).

## 9 **Materials and methods**

### 10 *Patient material*

11 In order to test the EORTC-GELA (Groupe d'Etude des Lymphomes de l'Adulte, presently known as  
12 the Lymphoma Study Association or LYSA) contouring guidelines for INRT as applied in the H10  
13 trial[6], two cases, typical of early stage HL, were selected. Case no 1 (male) had clinical stage (CS)  
14 IB disease in the left axilla and was treated with 4 cycles of ABVD (Adriamycin, Bleomycin,  
15 Vinblastine, Dacarbazine) followed by INRT. Case no 2 (female) had CS IIB disease in the  
16 mediastinum and bilateral neck and was treated with 6 cycles of ABVD followed by INRT.

17 The patients were scanned as recommended in the INRT guidelines[4]. Both patients were staged  
18 using whole body 18F-Fluorodeoxyglucose (FDG) PET/CT-scans performed before chemotherapy  
19 (from now referred to as "the pre-chemo PET/CT scan"). The pre-chemo PET/CT scans were  
20 acquired on a Siemens Biograph 40 (Siemens Healthineers, Erlangen, Germany) exactly 1 hour  
21 after injection of 400 MBq of FDG. Care was taken to acquire these images on a flat table top and  
22 in the same position as would later be used for the radiotherapy planning CT scan performed after  
23 chemotherapy (referred to as "the post-chemo CT scan"). The initial PET-positive volume was  
24 defined by visual evaluation of the FDG uptake, and was contoured on the pre-chemo PET/CT-scan  
25 by a nuclear medicine specialist as was standard practice in the host institution.

### 26 *Contouring process*

Contouring on the post-chemo CT scan and radiotherapy planning were carried out using the Eclipse® software from Varian Medical Systems (Palo Alto, USA). Scans and information were all anonymized and all contouring was performed at the institution where the patients were treated.

Eight individual radiation oncologists, each highly experienced in contouring for HL, participated in this study met at the host institution in order to contour on separate computers over the course of a single day and were blinded both to each other's contours and to the contours used for the patients' treatment. One of them (observer 2) contoured in collaboration with an experienced radiologist, as was standard practice in their institution. All observers contoured the initially involved volume on the pre-chemo PET/CT-scan with the help of the already contoured PET-positive volume. The decision as to which lymph nodes were involved initially was made on the basis of all available information including clinical and radiological information, the post-chemo CT-scan, and published guidelines [11]. The pre-chemo images were then fused with the post-chemo CT-scan. Each observer then modified the contours of the initially involved volume on the post-chemo CT-scan to allow for shrinkage of tissues from pre-to post-chemo scans, and to allow for uncertainties as deemed necessary. The resulting volume, defined as the tissue volume that contained the initially involved lymph nodes, was named the CTV and defined the tissue volume that each observer considered as needing irradiation.

#### *Assessment of the interobserver variability*

The CTV and planning target volume (PTV) from all observers were reported for each case. In order to facilitate the presentation of the data, a consensus contour was generated using an expectation-maximization algorithm for simultaneous truth and performance level estimation ("STAPLE"[12]) integrated in the publicly available research environment CERR[13] with a confidence level of 80% (chosen after visual evaluation). This algorithm has previously been used for the assessment of interobserver variation in radiotherapy [14,15]: in principle, the algorithm considers the whole collection of submitted CTV contours and generates a probabilistic estimate of the "true" CTV contour. The overlap between observer and consensus contour was investigated using the Sørensen-Dice similarity coefficient (DSC)[16,17], defined as:

$$DSC = \frac{2(A \cap B)}{A + B}$$

1 The magnitude of gross deviations between the surfaces of the observer and consensus contour  
2 was investigated using the Hausdorff distance [18] defined as:

3 “Hausdorff distance” ( $H_d$ ) or the maximum separation between two contours:  $H_d(X, Y) =$   
4  $\max_{x \in X} (\min_{y \in Y} d(x, y))$

5 In the context of radiotherapy, the Hausdorff distance can be thought of as reflecting the  
6 difference in beam aperture designed for different target volumes.

7 An ideal agreement between the surfaces of contours would then translate into a DSC of 1 and  
8 Hausdorff distance of 0 mm. All metrics were computed using the freeware 3D Slicer version 4.4  
9 ([www.slicer.org](http://www.slicer.org) [19]) and the extension SlicerRT[20].

10 The interobserver variation can be handled as a geometric uncertainty and included in the PTV  
11 margin using a margin recipe as a systematic geometric error[21]. In order to derive the systematic  
12 uncertainty resulting from the interobserver variation in this study, an in-house matlab script was  
13 designed to calculate the standard deviation (SD) between the surfaces of the observer contours  
14 in 6 directions (anterior, posterior, superior, inferior, right and left). Using the margin recipe  
15 described by van Herk et al [22], the SD can then be multiplied by 2.5 in order to provide a margin  
16 estimate accounting only for the interobserver variation (i.e. assuming other geometric  
17 uncertainties such as organ motion or patient set up are equal to 0).

18

### 19 *Treatment planning*

20 In more conformal techniques such as IMRT, interobserver variability may have a larger impact on  
21 the resulting dose distribution compared to simple forms of 3D-CRT (e.g., two opposing fields). In  
22 order to test this hypothesis, treatment plans were made based on each contour from the 8  
23 observers, using the current standard CTV-to-PTV margins recommended in the International  
24 Lymphoma Radiation Oncology Group (ILROG) guidelines[23]: a 1 cm isotropic margin was added,  
25 except for the case with mediastinal involvement (Case 2) where a 1.5 cm margin was added in the  
26 superior-inferior direction, as is recommended to account for respiration motion. All PTVs were  
27 then retracted 5 mm under the skin. For each set of contours a 3D-CRT plan and an intensity

1 modulated radiotherapy (IMRT) plan were generated. 3D-CRT plans often consisted of 2 opposing  
2 fields, with the addition of smaller fields to improve the dose homogeneity (“field-in-field”). IMRT  
3 plans used 4 to 5 different beam angles chosen to minimize entry through the organs at risk. The  
4 dose to the PTV was specified as 30.6 Gy in 17 fractions, 5 fractions per week. A total of 2x2x8  
5 treatment plans were calculated using the Analytic Anisotropic Algorithm (version 13, Varian  
6 Medical Systems). Similar target coverage was enforced for all plans, so that 95% of the PTV  
7 received at least 95% of the prescribed dose. The maximum dose accepted in the PTV (or any part  
8 of the body) was 107% of the prescribed dose. For IMRT plans, constraints were applied on the  
9 heart, lungs and female breasts (for Case 2) to achieve “as low as possible” a dose to those organs.  
10 The resulting mean doses to the heart and lungs, as well as the percentage of body volume  
11 receiving over 95% of the prescribed dose (V95%), were estimated for both cases. For case 2, the  
12 mean dose to the female breasts, thyroid and carotid artery were also reported.

13

## 14 **Results**

### 15 *Assessment of the interobserver variability*

16 The volumes of the CTVs defined by the eight observers and their associated PTVs are shown in  
17 tables 1 and 2. The median CTV was 120 cm<sup>3</sup> (IQR: 95-173 cm<sup>3</sup>) for case 1, and 255 cm<sup>3</sup> (IQR: 183-  
18 293 cm<sup>3</sup>). Compared to the consensus contours, this represented variations of -155% to 39% for  
19 case 1 and of -157% to 72% for case 2. These variations were carried on to the PTV volumes, with  
20 differences up to 268 cm<sup>3</sup> for case 1 and 366 cm<sup>3</sup> for case 2.

21 Representative slices with all observer contours as well the consensus contour are shown in  
22 figures 1 and 2. DSC values were generally high (>0.7), and the Hausdorff distances were around  
23 30 mm, with the notable exception of observer 8 (case 1), where the CTV was drawn as small  
24 “islands” (visible in figure 1). This configuration, however, did not lead to a smaller PTV size than  
25 for other observers: the PTV for observer 8 is very close to the median PTV (373 vs 389 cm<sup>3</sup>) even  
26 though the CTV was 5 times smaller than the median CTV.

1 The systematic uncertainty resulting from interobserver variation, expressed as the standard  
2 deviation between all observer contours in each direction, is reported in table S1 and ranged  
3 between 1.9 and 3.8 mm, with a median of 3.2 mm.

#### 4 *Impact on treatment planning*

5 All plans satisfied the PTV coverage criterion of 95% of the prescription dose to 95% of the PTV.  
6 The maximum dose allowed was kept under 107%, though this condition was difficult to fulfil for  
7 3D-CRT plans and small hot spots of up to 110% were occasionally accepted. The DVHs for the  
8 heart and lungs are presented in figures 3 and 4.

9 The resulting mean doses to the heart were low; around 0.3 Gy for case 1 and 0.9 Gy for case 2  
10 (see supplementary tables S2 and S3). For case 1, variations in mean dose were under 0.2 Gy for  
11 the heart and 1.2 Gy for the lungs across the group of observers, considering both treatment  
12 modalities. For case 2, mean heart dose variations were under 0.4 Gy with the noticeable  
13 exception of observer 7. There, the CTV encompassed more tissue in the inferior direction (figure  
14 2) and consequently led to a noticeably higher dose to the heart (for both modalities) and lungs  
15 (for IMRT) as observed by the DVHs in figure 4. The resulting mean heart dose was then increased  
16 to 4.4Gy (3D CRT) and 3.7 Gy (IMRT) in table S3. This increase was also present though less  
17 pronounced for the lungs and the breasts (table S4).

18

#### 19 **Discussion**

20 Inter- and intraobserver differences in the contouring of the gross tumour volume (GTV) and CTV  
21 have been reported in many tumour types and introduce a systematic geometrical uncertainty  
22 that must be taken into account in the subsequent planning process [22]. Quantifying this  
23 uncertainty is a difficult process and is usually performed using one or two representative patient  
24 examples [9,10,14]. Interobserver studies cannot claim to represent a whole clinical area or reflect  
25 the range of complexity of all patient cases. However, such studies offer a baseline for what is  
26 achievable, and a benchmark for other institutions developing their own treatment procedures  
27 and delineation guidelines. In our study, we chose to present two typical (as opposed to  
28 challenging) Hodgkin Lymphoma cases and to include only experts in the group of observers.



1 Combined to the close-to-optimal imaging conditions and INRT approach, we believe it represents  
2 a “best case scenario” situation which is substantially different from the two previously published  
3 reports of delineation uncertainty in HL [9,10].

4 Contouring for radiotherapy in the setting of modern combined modality treatment of early stage  
5 HL poses special challenges. In this situation, a volume which contained lymphoma before  
6 chemotherapy is contoured on a post-chemotherapy scan where most or all of the initial  
7 lymphoma has disappeared, and where shrinkage and deformation of the surrounding normal  
8 tissues has happened to varying degrees. This resembles in many ways the situation of post-  
9 operative radiotherapy in other tumour types, e.g. in head and neck cancer. In view of these  
10 challenges, it is reassuring to observe that the interobserver variability was on the same order of  
11 magnitude as has been reported for other indications, such as locally advanced lung cancer (4-5  
12 mm[24]), breast-conserving radiotherapy (2-8 mm, estimated from [25]) or prostate cancer (1.7-  
13 3.5 mm [22]).

14 It should, however, be noted that this level of confidence can only be achieved with dedicated  
15 radiation oncologists, knowledgeable of the disease, as well as an optimal use of imaging. It is  
16 strongly recommended to use PET as part of the pre-chemotherapy evaluation when planning to  
17 use highly conformal INRT after chemotherapy, as PET significantly improves the detection of  
18 involved sites in patients with HL[26,27]. It is also recommended that the pre-chemotherapy  
19 PET/CT-scan be acquired with the patient in the same position as will later be used for RT. Strict  
20 adherence to this principle is necessary if fusion of the pre-chemotherapy PET/CT images with the  
21 post-chemotherapy planning CT-images is to be used successfully. If this is not achievable, or if in  
22 spite of correct positioning the fusion of the pre-and post-chemotherapy images remains sub-  
23 optimal (e.g. if the patient loses weight), those additional uncertainties will mandate the use of  
24 larger margins to secure coverage of the initially-involved volume[23].

25 In this study, the interobserver variability led to marked differences in CTV volume. These  
26 differences are mitigated by the generation of the PTV margin, though large discrepancies remain.  
27 For example, for case 1, the PTV size for observer 1 is almost double that for observer 2. The  
28 impact of this variability on the dose to OARs appears modest, as the disease location was  
29 favourably far from major OARs in both cases. The mean doses to the heart, lung and breasts were

1 slightly higher with IMRT than with 3D-CRT but all remained very low. There is one notable  
2 exception: for case 2, observer 7, the mean heart dose was about 3.5 Gy higher than for all other  
3 observers, even though the corresponding PTV size, DSC and Hausdorff distance were all well  
4 within the reported range for other observers. This illustrates two things: 1) that all the mentioned  
5 metrics fail to fully represent the variability in doses received by the OARs and 2) that a few mm of  
6 difference in contouring can have a considerable impact on the dose to neighbouring OARs. This  
7 last difference will likely become even more substantial for novel treatment modalities such as  
8 proton therapy. Though a certain degree of variability is inevitable even within an expert group,  
9 efforts such as guidelines or “collaborative” contouring (with at least 2 observers present of the  
10 time of delineation) have been suggested to effectively decrease the risk of outliers and should be  
11 encouraged.

12 Determining the standard deviation between the observer contours is a first step towards  
13 calculating an evidence-based PTV margin for INRT in Hodgkin Lymphoma. The present results  
14 suggest that the systematic delineation uncertainty is around 3 mm, which alone would result in a  
15 CTV to PTV margin of almost 8 mm (according to the van Herk formula[22]), very close to the 10  
16 mm recommended by ILROG. In practice, other uncertainties, such as image fusion and patient  
17 set-up, must be included in a thorough margin recipe and can only increase the total PTV margin  
18 required. The limitations of our study include the small number of patient cases and relatively  
19 small number of observers. The clinical cases selected for this study, though fairly typical of early-  
20 stage HL, have smaller target volumes than the cases selected by Genovesi et al[9] and Piva et  
21 al[10], which could limit the comparison between these three studies in addition to the  
22 differences in imaging conditions already stated. Finally, all observers in this study were  
23 experienced with contouring for INRT for HL, and this might not reflect the experience of less  
24 experienced centres, or of centres using Involved Site radiotherapy (ISRT). Even bearing these  
25 limitations in mind, we believe that those results illustrate the potential of guidelines and  
26 standardization, both for pre- and post-chemotherapy imaging as well as for delineation of the  
27 CTV.

## 28 **Conclusions**

1 CTV volumes varied considerably between observers in both clinical cases. However, the  
2 systematic delineation uncertainty was around 3 mm and is comparable to that reported in other  
3 clinical situations. Results suggest that the dosimetric impact of interobserver variation is not  
4 larger for IMRT than for 3D-CRT. This study demonstrates that contouring target volumes for  
5 conformal INRT in HL can be performed with the same interobserver variability as can be achieved  
6 in other tumour types.

7

#### 8 **Disclosure of interest**

9 The authors have no conflict of interest regarding the data presented in this manuscript.

- 1 [1] Specht L, Gray RG, Clarke MJ, Peto R. Influence of more extensive radiotherapy and  
2 adjuvant chemotherapy on long-term outcome of early-stage Hodgkin's disease: A meta-  
3 analysis of 23 randomized trials involving 3,888 patients. *J Clin Oncol* 1998;16:830–43.
- 4 [2] Yahalom J, Mauch P. The involved field is back: issues in delineating the radiation field in  
5 Hodgkin's disease. *Ann Oncol* 2002;13:79–83. doi:10.1093/annonc/mdf616.
- 6 [3] Shahidi M, Kamangari N, Ashley S, Cunningham D, Horwich A. Site of relapse after  
7 chemotherapy alone for stage I and II Hodgkin's disease. *Radiother Oncol* 2006;78:1–5.  
8 doi:10.1016/j.radonc.2005.10.018.
- 9 [4] Girinsky T, van der Maazen R, Specht L, Aleman B, Poortmans P, Lievens Y, et al. Involved-  
10 node radiotherapy (INRT) in patients with early Hodgkin lymphoma: Concepts and  
11 guidelines. *Radiother Oncol* 2006;79:270–7. doi:10.1016/j.radonc.2006.05.015.
- 12 [5] Girinsky T, Ghalibafian M. Radiotherapy of Hodgkin Lymphoma: Indications, New Fields, and  
13 Techniques. *Semin Radiat Oncol* 2007;17:206–22. doi:10.1016/j.semradonc.2007.02.007.
- 14 [6] Raemaekers JMM, Andre MPE, Federico M, Girinsky T, Oumedaly R, Brusamolino E, et al.  
15 Omitting Radiotherapy in Early Positron Emission Tomography-Negative Stage I/II Hodgkin  
16 Lymphoma Is Associated With an Increased Risk of Early Relapse: Clinical Results of the  
17 Preplanned Interim Analysis of the Randomized EORTC/LYSA/FIL H10 Trial. *J Clin Oncol*  
18 2014;32:1188–94. doi:10.1200/JCO.2013.51.9298.
- 19 [7] Maraldo M V., Aznar MC, Vogelius IR, Petersen PM, Specht L. Involved Node Radiation  
20 Therapy: An Effective Alternative in Early-Stage Hodgkin Lymphoma. *Int J Radiat Oncol Biol*  
21 *Phys* 2013;85:1057–65. doi:10.1016/j.ijrobp.2012.08.041.
- 22 [8] Filippi AR, Ciammella P, Piva C, Ragona R, Botto B, Gavarotti P, et al. Involved-site image-  
23 guided intensity modulated versus 3d conformal radiation therapy in early stage  
24 supradiaphragmatic Hodgkin lymphoma. *Int J Radiat Oncol Biol Phys* 2014;89:370–5.  
25 doi:10.1016/j.ijrobp.2014.01.041.
- 26 [9] Genovesi D, Cèfaro GA, Vinciguerra A, Augurio A, Di Tommaso M, Marchese R, et al.  
27 Interobserver variability of clinical target volume delineation in supra-diaphragmatic  
28 Hodgkin's disease: a multi-institutional experience. *Strahlentherapie Und Onkol*  
29 2011;187:357–66. doi:10.1007/s00066-011-2221-y.
- 30 [10] Piva C, Genovesi D, Filippi AR, Balducci M, Barra S, Buglione M, et al. Interobserver  
31 variability in clinical target volume delineation for primary mediastinal B-cell lymphoma.  
32 *Pract Radiat Oncol* 2015;5:383–9. doi:10.1016/j.prro.2015.04.003.
- 33 [11] Girinsky T, Specht L, Ghalibafian M, Edeline V, Bonniaud G, Der Maazen R Van, et al. The  
34 conundrum of hodgkin lymphoma nodes: To be or not to be included in the involved node  
35 radiation fields. The EORTC-GELA lymphoma group guidelines. *Radiother Oncol*  
36 2008;88:202–10. doi:10.1016/j.radonc.2008.05.012.
- 37 [12] Warfield S, Zuh KH, Wells WM. Simultaneous Truth and Performance Level Estimation  
38 (STAPLE): An Algorithm for the Validation of Image Segmentation. *IEEE Trans Med Imaging*

2004;23:903–21. doi:10.1109/TMI.2004.828354.Simultaneous.

[13] Deasy JO, Blanco AI, Clark VH. CERR: a computational environment for radiotherapy research. *Med Phys* 2003;30:979–85. doi:10.1118/1.1568978.

[14] Michalski JM, Lawton C, El Naqa I, Ritter M, O’Meara E, Seider MJ, et al. Development of RTOG Consensus Guidelines for the Definition of the Clinical Target Volume for Postoperative Conformal Radiation Therapy for Prostate Cancer. *Int J Radiat Oncol Biol Phys* 2010;76:361–8. doi:10.1016/j.ijrobp.2009.02.006.

[15] Allozi R, Li XA, White J, Apte A, Tai A, Michalski JM, et al. Tools for Consensus Analysis of Experts’ Contours for Radiotherapy Structure Definitions. *Radiother Oncol* 2010;97:572–8. doi:10.1016/j.radonc.2010.06.009.Tools.

[16] Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology* 1945;26:297–302. doi:10.2307/1932409.

[17] Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol Skr* 1948;5:1–34.

[18] Hausdorff F. *Grundzuge der Mengenlehre*. Leipzig: Veit and Company; 1914.

[19] Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robbin J-C, Pujol S, et al. 3D Slicer as an image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging* 2012;30:1323–41. doi:10.1016/j.mri.2012.05.001.3D.

[20] Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G. SlicerRT: Radiation therapy research toolkit for 3D Slicer. *Med Phys* 2012;39:6332. doi:10.1118/1.4754659.

[21] Rasch C, Steenbakkers R, Van Herk M. Target definition in prostate, head, and neck. *Semin Radiat Oncol* 2005;15:136–45. doi:10.1016/j.semradonc.2005.01.005.

[22] van Herk M. Errors and margins in radiotherapy. *Semin Radiat Oncol* 2004;14:52–64. doi:10.1053/j.semradonc.2003.10.003.

[23] Specht L, Yahalom J, Illidge T, Berthelsen AK, Constine LS, Eich HT, et al. Modern Radiation Therapy for Hodgkin Lymphoma: Field and Dose Guidelines From the International Lymphoma Radiation Oncology Group (ILROG). *Int J Radiat Oncol* 2014;89:854–62. doi:10.1016/j.ijrobp.2013.05.005.

[24] Steenbakkers R, Duppen JC, Fitton I, Deurloo KEI, Zijp LJ, Comans EFI, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: A three-dimensional analysis. *Int J Radiat Oncol Biol Phys* 2006;64:435–48. doi:10.1016/j.ijrobp.2005.06.034.

[25] Hurkmans CW, Borger JH, Pieters BR, Russell NS, Jansen EPM, Mijnheer BJ. Variability in target volume delineation on CT scans of the breast. *Int J Radiat Oncol Biol Phys* 2001;50:1366–72. doi:10.1016/S0360-3016(01)01635-2.

[26] Girinsky T, Ghalibafian M, Bonniaud G, Bayla A, Magne N, Ferreira I, et al. Is FDG-PET scan in

1 patients with early stage Hodgkin lymphoma of any value in the implementation of the  
2 involved-node radiotherapy concept and dose painting? *Radiother Oncol* 2007;85:178–86.  
3 doi:10.1016/j.radonc.2007.07.003.

- 4 [27] Hutchings M, Loft A, Hansen M, Pedersen LM, Berthelsen AK, Keiding S, et al. Position  
5 emission tomography with or without computed tomography in the primary staging of  
6 Hodgkin's lymphoma. *Haematologica* 2006;91:482–9.

1

2 Figure legends

3 Figure 1: Delineations from the eight expert observers (in white) and generated consensus contour  
4 (in red) for case 1. Note the presence of a contour drawn as small “islands”, especially visible in  
5 the transverse view (top left).

6 Figure 2: Delineations from the eight expert observers (in white) and generated consensus contour  
7 (in red) for case 2. Note that one observer included more tissue in the inferior direction, towards  
8 the heart. This is especially visible in the sagittal view (bottom right).

9 Figure 3: DVH of the dose received by the heart and for all observer contours in Case 1. Results are  
10 shown for 3D conformal plans (left) as well as intensity-modulated radiotherapy plans (right).

11 Figure 4: DVH of the dose received by the heart and for all observer contours in Case 2. Results are  
12 shown for 3D conformal plans (left) as well as intensity-modulated radiotherapy plans (right).