

research article

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

Radiol Oncol 2016; 50(3): 254-262.

Received 29 December 2015

Accepted 1 February 2016

Correspondence to: Primož Petrič, M.D., M.Sc., Department of Radiation Oncology, National Centre for Cancer Care and Research, P.O. box 3050, Doha, Qatar. Phone: +974 44397720. E-mail: pp.visoko@gmail.com

Disclosure: No potential conflicts of interest were disclosed.

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 every day 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

Introduction

Modern radiotherapy techniques such as intensity modulated radiotherapy (IMRT), volumetric modulated arch therapy (VMAT) and image guided adaptive brachytherapy (IGABT) enable delivery of high doses to the target volume without escalating dose to organs at risk (OAR), offering the possibility of better local control while preserving good quality of life.^{1,2} Highly conformal radiation techniques and sharp dose falloff make the accuracy and precision of every step in treatment planning and delivery extremely important. Uncertainties in the process of radiotherapy include patient set-up error, inter- and intra-fraction organ movement, patient movement and uncer-

tainties in target volume delineation. Image guided radiation therapy (IGRT) addresses the uncertainties arising from patient set-up, patient and organ movement and improves target localisation during treatment. However, reduction of margins introduced with the use of IGRT is limited by the ability to adequately define the target. Accurate target volume delineation is a precondition for the use of IMRT, VMAT, IGABT and other high precision radiotherapy techniques, since all subsequent steps in treatment planning and delivery are based on target volume contours. Inadequate definition of the target introduces a systematic geographic miss that could potentially lead to reduction of the dose delivered to the tumour, lower local control and/or increased morbidity for an individual patient.³⁻⁶

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.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_{\max}/V_{\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 (κ) - statistics.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.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 and lung cancer, Hodgkin's lymphoma and sarcoma, where the V_{\max}/V_{\min} ratios are 6, 18.3, > 7, 15 and > 8, respectively.^{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.^{19,31-33} The ratio of V_{\max} to V_{\min} is 1.93 – 2.65 for GTV and 1.75 – 4.71 for CTV depending on imaging modal-

ity used for treatment planning.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_{\max}/V_{\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.⁶ For oesophageal cancer, median Jaccard conformity index for GTV was 0.69 in a study by Gwynne *et al.*³⁴, with the highest observer agreement in the middle section of the GTV, which is a marked improvement compared to results reported by Tai *et al.*¹⁸ at the start of 3D planning era, when V_{\max}/V_{\min} ratio was up to 6.

Cervix cancer

Considerable interobserver variability was described by Weiss *et al.*¹⁰ in CTV for cervix carcinoma, with the ratio of common to encompassing volume from 0.11 to 0.57 and V_{\max}/V_{\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.^{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_{\max}/V_{\min} ratio for CTV reported by Hong *et al.*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 *et al.*⁹ 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.²⁴

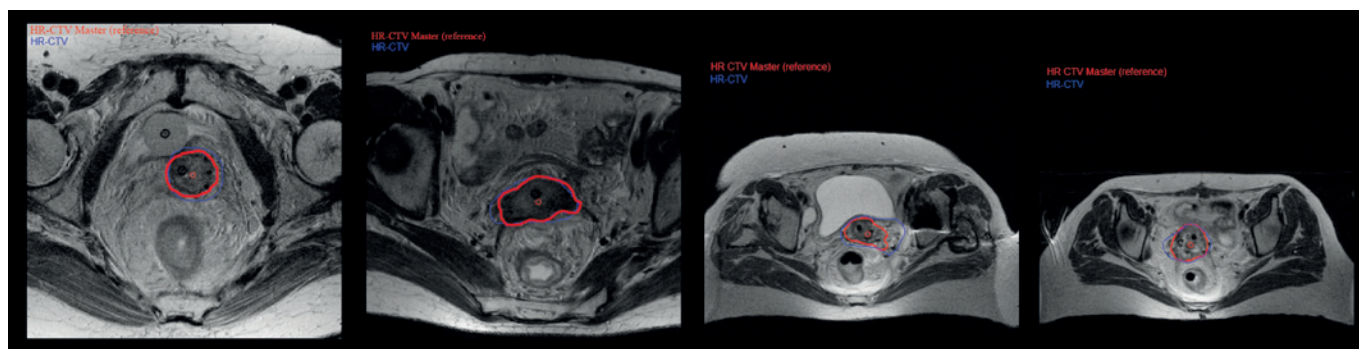


FIGURE 1. MR images showing interobserver variability between an unexperienced RO and the reference contour in IGABT of 4 cervix cancer patients (from a workshop for RO residents at the Institute of Oncology Ljubljana).

Lung tumours

In lung cancer the range of reported interobserver variability is quite large with V_{\max}/V_{\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 (*i.e.* atelectasis, peritumoral reaction) and normal structures (*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_{\max}/V_{\min} ratio from 1.3 to 2.8 and CI from 0.14 to 0.47 depending on imaging modality.^{16,37,38} 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_{\max}/V_{\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.^{39,40} 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_{\max}/V_{\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%.³ In cervix IGABT, a mean relative SD of 8-10% in D_{90} for GTV and HR CTV was observed in a single fraction analysis. For bladder and rectum mean relative SD for D_{2cc} was 5 – 8%, whereas for sigmoid it was 11%. When taking into account the whole treatment course, interobserver variability generated an uncertainty of +/-5 Gy ($\alpha\beta = 10$) for HRCTV and +/-2-3 Gy ($\alpha\beta = 3$) for OAR.⁴⁶

Strategies to improve target volume delineation

Several strategies to reduce uncertainties in target volume delineation have been proposed by different authors^{7,8,25} and there have been a few attempts to implement those strategies to improve quality assurance in clinical trials in radiation oncology.^{26,47-49} 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.⁵⁰

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

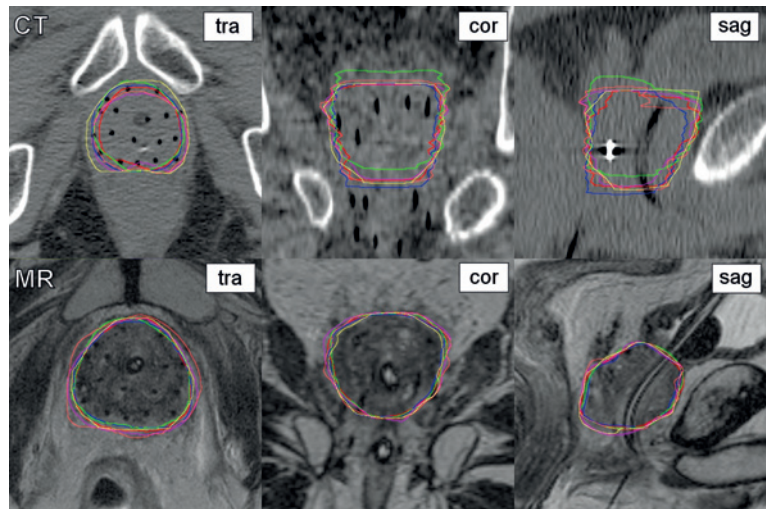


FIGURE 2. Interobserver variability in delineation of the prostate. MR and CT images in different planes of the same patient are shown. Ability to discriminate prostate apex, base and lateral borders is superior on MRI.

cor = coronal; sag = sagittal; tra = transverse⁴⁰

outside CT based GTV. The addition of MRI to CT did not result in significant improvement of CI.³¹ Patel *et al.*³³ also compared CT and PET-CT for delineation of primary and nodal GTV (GTV_p and GTV_n) in rectal cancer. Similarity index for GTV_p was modestly better, but statistically significant on PET CT *e.g.* 0.81 *vs.* 0.77, and notably better for GTV_n *e.g.* 0.70 *vs.* 0.22. Several studies show a good correlation between PET-CT and pathology based tumour length in oesophageal cancer^{51,52} but to our knowledge, there are no studies comparing interobserver variability on different imaging modalities. The benefit of PET-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 from 61% to 70% compared to CT alone.³⁶ 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.⁵³ 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.⁵⁴

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.^{9,24,55-57} In a study by Daisne *et al.*⁵⁵ 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_{CTPET} and GTV_{CTMR} was low, with CI = 0.62. When compared to the reference contour $CI_{CTPETMR}$ 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.*^{41,42} 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 V_{max} remaining up to 2.7 times larger than V_{min} .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.⁶¹⁻⁶⁷ 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 study⁶⁸, 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.⁶⁹ 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

TABLE 1. Interobserver variation for various tumour sites

Tumoursite	Target volume	No of pts	No of obs	Imaging modality	Results	Author (publication date)
Rectum	GTV, CTV	2	10	CT, PETCT	CI(GTV) = 0.26-0.33 CI(CTV) = 0.29-0.35	Krengli et al 2010
	GTV	52	5	CT, PETCT, MRI	CI = 0.79-0.93	Buijssen et al 2012
	CTV	8	10	CT	CI = 0.63-0.66	Nijkamp et al 2012
	GTV	6	4	CT, PETCT	SI(GTV-P) = 0.77-0.81 SI(GTV-N) = 0.22-0.70	Patel et al 2007
Stomach	CTV, PTV	1	10	CT	Vmax/Vmin(CTV) = 3.4 Vmax/Vmin(PTV) = 2.6	Jansen et al 2010
Oesophagus	GTV	1	50	CT	JCI = 0.69	Gwynne et al 2012
	GTV, CTV, PTV	1	48	CT	Vmax/Vmin(PTV) = 5.25-6.03	Tai et al 1998
Cervix EBRT	CTV	3	7	CT	CI = 0.11-0.57 Vmax/Vmin = 3.6-4.9	Weiss et al 2003
IGABT	GTV, HRCTV, IRCTV	6	10	MRI	CI(GTV) = 0.6-0.8 CI(HR&IRCTV) = 0.6-0.7	Petrič et al 2012, 2013
Head and neck	GTV,CTV,PTV	1	20	CT	Vmax/Vmin(CTV) = 18.3	Hong et al 2012
	GTV	41	3	CT, PETCT, MRI	CI(GTV-P) = 0.54-0.62 CI(GTV-N)>0.75	Thiagajaran et al 2012
	CTV, CTVE	10	10	CT, MRI	localSD(CTV) = 3.3-4.4mm localSD(CTVe) = 4.9-5.9mm	Rasch et al 2012
Lung	GTV	12	8	CT, CBCT	CI = 0.27-0.39 CIgen = 0.58-0.70	Altorjai et al 2012
	GTV	8	5	CT	Vmax/Vmin>7	Van De Steene et al 2002
	GTV	10	17	CT	Vmax/Vmin = 5.2 CI = 0.04-0.48	Giraud et al 2002
	GTV	22	11	CT, PETCT	meanCI = 0.17(CT),0.29(PETCT) localSD = 1cm(CT),0.4cm(PETCT)	Steenbakkers et al 2006
	GTV	19	2	CT, PETCT	medianCI(CT) = 0.61, medianCI(PETCT) = 0.70	Fox et al 2005
Brain	CTV	7	5	CT + MRI	CI = 0.14-0.47	Cattaneo et al 2005
	GTV	5	9	CT, MRI	Vmax/Vmin(CT) = 1.7-2.8 Vmax/Vmin(MR) = 1.5-2.7	Weltens et al 2001
Prostate	Prostate, seminal vesicles (SV)	10	7	CT	Vmax/Vmin(P) = 1.18-1.63 Vmax/Vmin(SV) = 2.02-6.43	Valicenti et al 1999
	Prostate	3	2	CT	Vmax/Vmin = 1.39-1.65	Seddon et al 2000
	Prostate	5	5	CT, MRI	MeanCI(MR)CI = 0.83 MeanCI(CT) = 0.69	Segedin et al 2011
Breast	Lumpectomy cavity (LC), CTV	15	3	CT, MRI	CI(LC) = 0.32(MR),0.52(CT) CI(CTV) = 0.77(CT),0.79(MR)	Giezen et al 2011,2012
	Lumpectomy cavity	30	5	CT	MeanCI = 0.36	Boersma et al 2012
	Lumpectomy cavity, CTV, PTV	8	13	CT	CI(LC) = 0.19-0.77 CI(CTV) = 0.38-0.80 CI(PTV) = 0.45-0.81	VanMourik et al 2010
	Lumpectomy cavity, PTV	9 5	5 4	CT	CI(LC) = 0.48-0.52 CI(PTV) = 0.55-0.59 Vmax/Vmin = 2.2-2.8	Major et al 2015
	Lumpectomy cavity, CTV	18	5	CT	MeanCI(LC) = 0.56 MeanCI(CTV) = 0.87	Struikmans et al 2005

CI = 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;

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.⁴³ 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.⁷¹ Mean V_{max}/V_{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.⁷³ 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.^{15,37,74} While Hurkmans *et al.*⁷⁴ reported that more experienced ROs delineate smaller volumes than unexperienced in breast carcinomas, Giraud *et al.*¹⁵ 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.³⁷ Only a few publications have addressed the subject of training, some in the course of pre-accrual quality assurance delineation exercises (dummy run).^{26,34,47-49,75,76} 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).⁴⁸ 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).⁴⁹ Dewas *et al.*⁷⁵, 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, V_{20} 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 at-

tention needs to be paid to predilection areas for larger interobserver variability, identified in available literature.^{25,26,30,39,40}

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 multimodality 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

1. Pötter R, Georg P, Dimopoulos JC, Grimm M, Berger D, Nesvacil N, et al. Clinical outcome of protocol based image (MRI) guided adaptive brachytherapy combined with 3D conformal radiotherapy with or without chemotherapy in patients with locally advanced cervical cancer. *Radiother Oncol* 2011; **100**: 116-23.

2. van Rij C, Oughlane-Heemsbergen W, Akerstaff A, Lamers E, Balm A, Rasch C. Parotid gland sparing IMRT for head and neck cancer improves xerostomia related quality of life. *Radiat Oncol* 2008; **3**: 41.
3. Van de Steene J, Linthout N, De Mey J, Vinh-Hung V, Claessens C, Noppen M, et al. Definition of gross tumor volume in lung cancer: Inter-observer variability. *Radiother Oncol* 2002; **62**: 37-9.
4. Lee WR, Roach M. III, Michalski J, Moran B, Beyer D. Interobserver variability leads to significant differences in quantifiers of prostate implant adequacy. *Int J Radiat Oncol Biol Phys* 2002; **54**: 457-61.
5. Kim RY, McGinnis SL, Spencer SA, Meredith RF, Jennelle RLS, Salter MM. Conventional four-field pelvic radiotherapy technique without computed tomography-treatment planning in cancer of the cervix: potential geographic miss and its impact on pelvic control. *Int J Radiat Oncol Biol Phys* 1995; **31**: 109-12.
6. Jansen EPM, Nijkamp J, Gubanski M, Lind PARM, Verheij M. Interobserver variation of clinical target volume delineation in gastric cancer. *Int J Radiat Oncol* 2010; **77**: 1166-70.
7. Njeh CF, Dong L, Orton CG. IGRT has limited clinical value due to lack of accurate tumor delineation. *Med Phys* 2008; **33**: 136-40.
8. Weiss E, Hess CF. The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy. *Strahlenther Onkol* 2003; **179**: 21-30.
9. Thiagarajan A, Caria N, Schöder H, Iyer NG, Wolden S, Wong RJ, et al. Target volume delineation in oropharyngeal cancer: impact of PET, MRI, and physical examination. *Int J Radiat Oncol* 2012; **83**: 220-7.
10. Weiss E, Richter S, Krauss T, Metzkelin SI, Hille A, Pradier O, et al. Conformal radiotherapy planning of cervix carcinoma: differences in the delineation of the clinical target volume. A comparison between gynaecologic and radiation oncologists. *Radiother Oncol* 2003; **67**: 87-95.
11. Boersma LJ, Janssen T, Elkhuizen PHM, Poortmans P, van der Sangen M, Scholten AN, et al. Reducing interobserver variation of boost-CTV delineation in breast conserving radiation therapy using a pre-operative CT and delineation guidelines. *Radiother Oncol* 2012; **103**: 178-82.
12. van Mourik AM, Elkhuizen PHM, Minkema D, Duppen JC, van Vliet-Vroegindewij C. Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines. *Radiother Oncol* 2010; **94**: 286-91.
13. Struikmans H, Wárlám-Rodenhuis C, Stam T, Stapper G, Tersteeg RJ, Bol GH, et al. Interobserver variability of clinical target volume delineation of glandular breast tissue and of boost volume in tangential breast irradiation. *Radiother Oncol* 2005; **76**: 293-9.
14. Altorjai G, Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D, et al. Cone-beam CT-based delineation of stereotactic lung targets: the influence of image modality and target size on interobserver variability. *Int J Radiat Oncol* 2012; **82**: e265-e72.
15. Giraud P, Elles S, Helfre S, De Rycke Y, Servois V, Carette MF, et al. Conformal radiotherapy for lung cancer: Different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. *Radiother Oncol* 2002; **62**: 27-36.
16. Cattaneo GM, Reni M, Rizzo G, Castellone P, Ceresoli GL, Cozzarini C, et al. Target delineation in post-operative radiotherapy of brain gliomas: interobserver variability and impact of image registration of MR(pre-operative) images on treatment planning CT scans. *Radiother Oncol* 2005; **75**: 217-23.
17. Rasch C, Steenbakkers R, van Herk M. Target definition in prostate, head, and neck. *Semin Radiat Oncol* 2005; **15**: 136-45.
18. Tai P, Van Dyk J, Yu E, Battista J, Stitt L, Coad T. Variability of target volume delineation in cervical esophageal cancer. *Int J Radiat Oncol Biol Phys* 1998; **42**: 277-88.
19. Krengli M, Canillo B, Turri L, Bagnasacco P, Berretta L, Ferrara T, et al. Target volume delineation for preoperative radiotherapy of rectal cancer: interobserver variability and potential impact of FDG-PET/CT imaging. *Technol Cancer Res Treat* 2010; **9**: 393-8.
20. Hamilton CS, Joseph DJ, Skov A, Denham JW. CT scanning for definitive radiotherapy planning of prostate cancer: necessity or nicety? Results from survey of radiation oncologists working at different institutions in Australasia. *Australas Radiol* 1990; **34**: 288-92.
21. Valicenti RK, Sweet JW, Hauck WW, Hudes RS, Lee T, Dicker AP, et al. Variation of clinical target volume definition in three-dimensional conformal radiation therapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 1999; **44**: 931-5.
22. Tanderup K, Nesvacil N, Pötter R, Kirisits C. Uncertainties in image guided adaptive cervix cancer brachytherapy: Impact on planning and prescription. *Radiother Oncol* 2013; **107**: 1-5.
23. Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. *Strahlenther Onkol* 2012; **188**: 160-7.
24. Rasch CRN, Steenbakkers RJHM, Fitton I, Duppen JC, Nowak PJ, Pameijer FA, et al. Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer. *Radiat Oncol* 2010; **5**: 21.
25. Petrič P, Hudej R, Rogelj P, Blas M, Tanderup K, Fidarova E, et al. Uncertainties of target volume delineation in MRI guided adaptive brachytherapy of cervix cancer: a multi-institutional study. *Radiother Oncol* 2013; **107**: 6-12.
26. Seddon B, Bidmead M, Wilson J, Khoo V, Dearnaley D. Target volume definition in conformal radiotherapy for prostate cancer: quality assurance in the MRC RT-01 trial. *Radiother Oncol* 2000; **56**: 73-83.
27. Steenbakkers RJHM, Duppen JC, Fitton I, Deurloo KE, Zijp LJ, Comans EF, 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.
28. Genovesi D, Céfaro GA, Vinciguerra A, Augurio A, Di Tommaso M, Marchese R, et al. Interobserver variability of clinical target volume delineation in supra-diaphragmatic (Hodgkin)'s disease: a multi-institutional experience. *Strahlenther Onkol* 2011; **187**: 357-66.
29. Genovesi D, Ausili Céfaro G, Trignani M, Vinciguerra A, Augurio A, Di Tommaso M, et al. Interobserver variability of clinical target volume delineation in soft-tissue sarcomas. *Cancer Radiother* 2014; **18**: 89-96.
30. Hong TS, Tome WA, Harari PM. Heterogeneity in head and neck IMRT target design and clinical practice. *Radiother Oncol* 2012; **103**: 92-8.
31. Buijssen J, van den Bogaard J, van der Weide H, Engelsman S, van Stiphout R, Janssen M, et al. FDG-PET-CT reduces the interobserver variability in rectal tumor delineation. *Radiother Oncol* 2012; **102**: 371-6.
32. Nijkamp J, de Haas-Kock DFM, Beukema JC, Neelis KJ, Woutersen D, Ceha H, et al. Target volume delineation variation in radiotherapy for early stage rectal cancer in the Netherlands. *Radiother Oncol* 2012; **102**: 14-21.
33. Patel DA, Chang ST, Goodman KA, Quon A, Thorndyke B, Gambhir SS, et al. Impact of integrated PET/CT on variability of target volume delineation in rectal cancer. *Technol Cancer Res Treat* 2007; **6**: 31-6.
34. Gwynne S, Spezi E, Wills L, Nixon L, Hurt C, Joseph G, et al. Toward semi-automated assessment of target volume delineation in radiotherapy trials: The SCOPE 1 Pretrial Test Case. *Int J Radiat Oncol* 2012; **84**: 1037-42.
35. Petric P, Hudej R, Rogelj P, Blas M, Segedin B, Logar HB, et al. Comparison of 3D MRI with high sampling efficiency and 2D multiplanar MRI for contouring in cervix cancer brachytherapy. *Radiol Oncol* 2012; **46**: 242-51.
36. Fox JL, Rengan R, O'Meara W, Yorke E, Erdi Y, Nehmeh S, et al. Does registration of PET and planning CT images decrease interobserver and intraobserver variation in delineating tumor volumes for non-small-cell lung cancer? *Int J Radiat Oncol Biol Phys* 2005; **62**: 70-5.
37. Leunens G, Weltens A, Vertraete J, van der Schuren E. Quality assessment of medical decision making in radiation oncology: variability in target volume delineation for brain tumours. *Radiother Oncol* 1993; **29**: 169-75.
38. Weltens C, Menten J, Feron M, Bellon E, Demaerel P, Maes F, et al. Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging. *Radiother Oncol* 2001; **60**: 49-59.
39. Villeirs GM, Van Vaerenbergh K, Vakaet L, Bral S, Claus F, De Neve WJ, et al. Interobserver delineation variation using CT versus combined CT + MRI in intensity-modulated radiotherapy for prostate cancer. *Strahlenther Onkol* 2005; **181**: 424-30.
40. Segedin B, But Hadzic J, Rogelj P, Sesek M, Zobec Logar HB, Kragelj B, et al. Interobserver variation in MRI and CT based contouring for prostate cancer. [Abstract]. *Radiother Oncol* 2011; **99(Suppl 1)**: S285.

41. Giezen M, Kouwenhoven E, Scholten AN, Coerkamp EG, Heijnen M, Jansen WP, et al. MRI- versus CT-based volume delineation of lumpectomy cavity in supine position in breast-conserving therapy: An exploratory study. *Int J Radiat Oncol Biol Phys* 2012; **82**: 1332-40.
42. Giezen M, Kouwenhoven E, Scholten AN, Coerkamp EG, Heijnen M, Jansen WP, et al. Magnetic resonance imaging- versus computed tomography-based target volume delineation of the glandular breast tissue (clinical target volume breast) in breast-conserving therapy: an exploratory study. *Int J Radiat Oncol* 2011; **81**: 804-11.
43. Major T, Gutiérrez C, Guix B, Móza E, Hannoun-Levi JM, Lössl K, et al. Interobserver variations of target volume delineation in multicatheter partial breast brachytherapy after open cavity surgery. *Brachytherapy* 2015; **14**: 925-32.
44. Steenbakkers RJH, Deurloo KE, Nowak PJ, Lebesque JV, van Herk M, Rasch CR. Reduction of dose delivered to the rectum and bulb of the penis using MRI delineation for radiotherapy of the prostate. *Int J Radiat Oncol* 2003; **57**: 1269-79.
45. Emami B, Sethi A, Petruzzelli GJ. Influence of MRI on target volume delineation and IMRT planning in nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys* 2003; **57**: 481-8.
46. Hellebust TP, Tanderup K, Lervåg C, Fidarova E, Berger D, Malinen E, et al. Dosimetric impact of interobserver variability in MRI-based delineation for cervical cancer brachytherapy. *Radiother Oncol* 2013; **107**: 13-9.
47. Gwynne S, Spezi E, Sebag-Montefiore D, Mukherjee S, Miles E, Conibear J, et al. Improving radiotherapy quality assurance in clinical trials: assessment of target volume delineation of the pre-accrual benchmark case. *Br J Radiol* 2013; **86**: 20120398.
48. Schimek-Jasch T, Troost EG, Rucker G, Prokic V, Avlar M, Duncker-Rohr V, et al. A teaching intervention in a contouring dummy run improved target volume delineation in locally advanced non-small cell lung cancer: Reducing the interobserver variability in multicentre clinical studies. *Strahlenther Onkol* 2015; **191**: 525-33.
49. Khoo ELH, Schick K, Plank AW, Poulsen M, Wong WW, Middleton M, et al. Prostate contouring variation: Can it be fixed? *Int J Radiat Oncol Biol Phys* 2012; **82**: 1923-29.
50. Steenbakkers RJHM, Duppen JC, Fitton I, Deurloo KE, Zijl L, Uitterhoeve AL, et al. Observer variation in target volume delineation of lung cancer related to radiation oncologist-computer interaction: A "Big Brother" evaluation. *Radiother Oncol* 2005; **77**: 182-90.
51. Mamede M, El Fakhri G, Abreu-e-Lima P, Gandler W, Nosé V, Gerbaudo VH. Pre-operative estimation of esophageal tumor metabolic length in FDG-PET images with surgical pathology confirmation. *Ann Nucl Med* 2007; **21**: 553-62.
52. Zhong X, Yu J, Zhang B, Mu D, Zhang W, Li D, et al. Using 18F-fluorodeoxyglucose positron emission tomography to estimate the length of gross tumor in patients with squamous cell carcinoma of the esophagus. *Int J Radiat Oncol Biol Phys* 2009; **73**: 136-41.
53. Bradley J, Bae K, Choi N, Forster K, Siegel BA, Brunetti J, et al. A phase II comparative study of gross tumor volume definition with or without PET/CT fusion in dosimetric planning for non-small-cell lung cancer (NSCLC): primary analysis of Radiation Therapy Oncology Group (RTOG) 0515. *Int J Radiat Oncol Biol Phys* 2012; **82**: 435-41.
54. Schaefer A, Kim YJ, Kremp S, Mai S, Fleckenstein J, Bohnenberger H, et al. PET-based delineation of tumor volumes in lung cancer: comparison with pathological findings. *Eur J Nucl Med Mol Imaging* 2013; **40**: 1233-44.
55. Daisne J-F, Duprez T, Weynand B, Lonnew M, Hamoir M, Reyckler H, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology* 2004; **233**: 93-100.
56. Breen SL, Publicover J, De Silva S, Pond G, Brock K, O'Sullivan B, et al. Intraobserver and Interobserver Variability in GTV Delineation on FDG-PET-CT Images of Head and Neck Cancers. *Int J Radiat Oncol* 2007; **68**: 763-70.
57. Geets X, Daisne J-F, Arcangeli S, Coche E, De Poel M, Duprez T, et al. Interobserver variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI. *Radiother Oncol* 2005; **77**: 25-31.
58. Anderson CM, Sun W, Buatti JM, Maley JE, Policeni B, Mott SL, et al. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR images of head and neck cancer. *J Radiat Oncol* 2014; **1**: 006.
59. Rasch CRN, Barillot I, Remeijer P, Touw A, van Herk M, Lebesque J V. Definition of the prostate in CT and MRI: a multi-observer study. *Int J Radiat Oncol Biol Phys* 1999; **43**: 57-66.
60. Barkati M, Simard D, Taussky D, Delouya G. Magnetic resonance imaging for prostate bed radiotherapy planning: An inter- and intra-observer variability study. *J Med Imaging Radiat Oncol* 2015; Nov 16. doi: 10.1111/1754-9485.12416. [Epub ahead of print].
61. Haie-Meder C, Pötter R, Van Limbergen E, Briot E, De Brabandere M, Dimopoulos J, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (I): concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV. *Radiother Oncol* 2005; **74**: 235-45.
62. Small W Jr, Mell LK, Anderson P, Creutzberg C, De Los Santos J, Gaffney D, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy in postoperative treatment of endometrial and cervical cancer. *Int J Radiat Oncol Biol Phys* 2008; **71**: 428-34.
63. Harris VA, Staffurth J, Naismith O, Esmail A, Gulliford S, Khoo V et al. Consensus guidelines and contouring atlas for pelvic node delineation in prostate and pelvic node intensity modulated radiation therapy. *Int J Radiat Oncol* 2015; **92**: 874-83.
64. Grégoire V, Ang K, Budach W, Grau C, Hamoir M, Langendijk JA, et al. Delineation of the neck node levels for head and neck tumors: A 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. *Radiother Oncol* 2014; **110**: 172-81.
65. Nielsen MH, Berg M, Pedersen AN, Andersen K, Glavicic V, Jakobsen EH, et al. Delineation of target volumes and organs at risk in adjuvant radiotherapy of early breast cancer: National guidelines and contouring atlas by the Danish Breast Cancer Cooperative Group. *Acta Oncol (Madr)* 2013; **52**: 703-10.
66. Lim K, Small W, Portelance L, Creutzberg C, Jürgenliemk-Schulz IM, Mundt A, et al. Consensus guidelines for delineation of clinical target volume for intensity-modulated pelvic radiotherapy for the definitive treatment of cervix cancer. *Int J Radiat Oncol* 2011; **79**: 348-55.
67. Wu AJ, Bosch WR, Chang DT, Hong TS, Jabbour SK, Kleinberg LR et al. Expert consensus contouring guidelines for intensity modulated radiation therapy in esophageal and gastroesophageal junction cancer. *Int J Radiat Oncol* 2015; **92**: 911-20.
68. Fuller CD, Nijkamp J, Duppen J, Rasch CR, Thomas CR Jr, Wang SJ, et al. Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting. *Int J Radiat Oncol Biol Phys* 2011; **79**: 481-9.
69. Wong EK, Truong PT, Kader HA, Nichol AM, Salter L, Petersen R, et al. Consistency in seroma contouring for partial breast radiotherapy: impact of guidelines. *Int J Radiat Oncol Biol Phys* 2006; **66**: 372-6.
70. Senan S, van Sornsens de Koste J, Samson M, Tankink H, Jansen P, Nowak PJ, et al. Evaluation of a target contouring protocol for 3D conformal radiotherapy in non-small cell lung cancer. *Radiother Oncol* 1999; **53**: 247-55.
71. Mitchell DM, Perry L, Smith S, Elliott T, Wylie JP, Cowan RA, et al. Assessing the effect of a contouring protocol on postprostatectomy radiotherapy clinical target volumes and interphysician variation. *Int J Radiat Oncol* 2009; **75**: 990-3.
72. Bowden P, Fisher R, Mac Manus M, Wirth A, Duchesne G, Millward M, et al. Measurement of lung tumor volumes using three-dimensional computer planning software. *Int J Radiat Oncol Biol Phys* 2002; **53**: 566-73.
73. Jefferies S, Taylor A, Reznick R. Results of a national survey of radiotherapy planning and delivery in the UK in 2007. *Clin Oncol* 2009; **21**: 204-17.
74. 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.
75. Dewas S, Bibault J-E, Blanchard P, Vautravers-Dewas C, Pointreau Y, Denis F, et al. Delineation in thoracic oncology: a prospective study of the effect of training on contour variability and dosimetric consequences. *Radiat Oncol* 2011; **6**: 118.
76. Szumacher E, Harnett N, Warner S, Kelly V, Danjoux C, Barker R, et al. Effectiveness of educational intervention on the congruence of prostate and rectal contouring as compared with a gold standard in three-dimensional radiotherapy for prostate. *Int J Radiat Oncol Biol Phys* 2010; **76**: 379-85.