Radiotherapy and Oncology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original article

Variability of clinical target volume delineation for definitive radiotherapy in cervix cancer

Gemma Eminowicz *, Mary McCormack

University College Hospital, London, UK

ARTICLE INFO

Article history:
Received 11 May 2015
Received in revised form 7 October 2015
Accepted 8 October 2015
Available online xxxx

Keywords:
Inter-observer variation
Cervical cancer radiotherapy
Radiotherapy delineation

ABSTRACT

Background/purpose: Accurate target volume delineation is essential for radiotherapy delivery, yet significant intra and inter-observer variability is documented. We analysed the variation in cervical cancer clinical target volume (CTV) delineation.

Materials/methods: All INTERLACE participating centres completed two RTQA outlining exercises. The Trial Management Group created a consensus outline. A separate STAPLE algorithm outline was created. Using these two outlines an optimised gold standard was generated. Volume, maximum distance from DICOM centre in all directions, and Jaccard Conformity Index (JCI) were calculated and compared for each centres' outlines. Anatomical areas included within CTV were recorded to detect systematic differences.

Results: 21 outlines were compared for case 1 and 22 for case 2. Volume ranged from 340 cc to 676 cc (case 1) and from 458 cc to 806 cc (case 2). A maximum 4 cm difference between outlines was observed in one direction. JCI ranged from 0.51 to 0.81 (case 1) and 0.57 to 0.81 (case 2). Variation in anatomical areas included in CTV exists between the two cases and between centres.

Conclusions: Significant inter-observer variation in cervical cancer delineation has been demonstrated. Ongoing efforts are needed to ensure inter-observer consistency through education, guidelines and multi-centre collaboration.

 Radiotherapy and Oncology xxx (2015) xxx–xxx

Many centres within Europe and the UK use Intensity Modulated Radiotherapy (IMRT) for the curative treatment of cervical cancer. IMRT reduces dose delivered to organs at risk (OAR) compared with 3D conformal radiotherapy [1,2] leading to reduced toxicity rates [3,4]. To ensure precise target coverage when using IMRT it is vital to delineate the gross tumour volume (GTV), clinical target volume (CTV) and OARs accurately. However, one of the largest uncertainties within radiotherapy planning is target volume delineation. Significant inter-observer variability exists across many tumour sites, including oesophageal, prostate, head and neck, bladder, breast and lung [5–10]. Detailed guidelines and systematic training can reduce variation in GTV and CTV delineation in image guided cervical brachytherapy [11,12]. Only a few published papers assess variation of external beam radiotherapy (EBRT) GTV and CTV delineation for cervical cancer [13,14]. No published papers review delineation standards across many UK centres.

The INTERLACE trial is a phase III multicenter trial of weekly induction chemotherapy followed by chemo-radiation versus chemo-radiation alone in locally advanced cervical cancer. As part of the INTERLACE radiotherapy quality assurance (RTQA) participating centres outline and plan two test cases following trial protocol. This paper is the first to analyse cervical EBRT outlining variation in detail by comparing these RTQA cases between 21 UK INTERLACE centres.

Materials/methods

Delineation process

To complete the INTERLACE RTQA each investigator delineated two test cases (Appendix 1); FIGO 3B (right side-wall) squamous carcinoma (case 1); FIGO 2B squamous carcinoma with lower uterine involvement (case 2). All investigators could access the RTQA pack (Appendix 1) including the anonymised patient history, examination-under-anaesthetic (EUA) findings and imaging reports. The diagnostic MRI images were not available. Standard UK practice is CT outlining using diagnostic MRI and EUA for reference. The planning CT (0.25 cm slice thickness), available in DICOM format, could be imported into any treatment planning system (TPS). This allowed investigators to use the TPS which they use daily. The first 10 centres completed delineation using INTERLACE protocol version 1 (V1) and the subsequent 11 centres used an

* Corresponding author at: 250 Euston Road, London, NW1 2PG, UK.
E-mail address: gemmaeminowicz@nhs.net (G. Eminowicz).

http://dx.doi.org/10.1016/j.radonc.2015.10.007
0167-8140/

updated version (V2). The protocols define two independent CTVs; primary CTV (CTV1) and nodal CTV (CTV2). V1 and V2 differ regarding inclusion of upper vagina and bilateral parametria in CTV2 for V1 and CTV1 for V2. No other guidance differs between V1 and V2. Therefore, both protocols recommend inclusion of the same anatomical areas within the combined CTV1 and CTV2: tumour, entire cervix, bilateral parametria, ovaries if seen, upper vagina, entire uterus, and pelvic nodal areas discussed later. Delineation review by the INTERLACE RTQA team was performed. All cases that were not protocol compliant were edited and resubmitted and were approved once fully protocol compliant.

Gold standard delineation

5 experienced clinicians from different UK centres, who are members of the trial management group (TMG), downloaded the test cases and delineated them following protocol. These independent outlines were collated in DICOM format and were visually and quantitatively simultaneously reviewed by the TMG. A consensus outline for each OAR and CTV was then manually created. This is the 'TMG gold standard'.

A separate 'simultaneous truth and performance level estimation' (STAPLE), described by Warfield et al., was created for both test cases from all centres' cases [15]. The STAPLE algorithm applies an expectation–maximisation algorithm to multiple outlines of one case to compute a probabilistic estimate of the true (gold standard) outline. STAPLE weights each outline on the estimated performance level and incorporates spatial distribution and homogeneity constraints models [15]. To create the STAPLE, an outline agreement confidence level must be selected. For our algorithm we applied 95% confidence level. The TMG gold standard was validated using this STAPLE outline. Areas of variation between the TMG and STAPLE outline were reviewed and an 'optimised gold standard' outline (GSCTV1 + 2) was generated.

Delineation comparison

'CTV1 + 2' was analysed to allow comparison of all cases together. Individual analysis of CTV1 and CTV2 would highlight inconsistencies due to protocol variation in vagina and parametrial inclusion. Anatomical areas included in CTV1 + 2 for V1 and V2 are detailed earlier. Each CTV1 + 2 outline was imported into CERR (computational environment for radiotherapy research software) [16] and SHERRI (surrey heuristic engine for radiotherapy radiobiology and imaging). Maximum distance from DICOM centre (i.e. CT reference point), volume (CTVI + 2), Jaccard conformity index (JCI) and anatomical regions included were analysed on CERR. Volume and JCI results were validated using SHERRI.

Maximum distance from the DICOM centre in all 6 directions was documented. This was calculated by recording the most extreme *X*, *Y* or *Z* coordinate in all 6 directions (anterior, posterior, inferior, superior, left and right lateral) on which CTV1 + 2 is seen. The *X*, *Y* and *Z* coordinates represent the distance in centimetres from the DICOM centre. These distances are not necessarily at the same point along the axis. The most extreme point in one direction for two independent outlines may therefore be in different anatomical regions.

CTV1 + 2 total volume was calculated on CERR and SHERRI. The average was recorded as the result.

The JCI was calculated for each outline against GSCTV1 + 2. JCI is the ratio of common volume to encompassed volume. It is calculated using JCI = A∩B/AUB i.e. intersection volume of A and B ÷ union volume of A and B. For perfectly overlapping outlines JCI = 1.0, JCI = 0.5 equates to a 66% overlap, JCI = 0.6 75% overlap, 0.7 approximately 82% overlap and 0.82 90% overlap. This calculation was performed manually on CERR validating the programmed SHERRI calculation. The optimal JCI result is unclear from the literature. A poor outline correlation is represented by JCI <0.5 [17,18]. Gwynne et al. suggested JCI $\geqslant$0.7 is acceptable [19]. This is the level we applied.

Each outline was visually reviewed on CERR. The proportion of outlines which included the following anatomical areas was recorded:

- Common iliac nodal region,
- Internal iliac nodal region,
- External iliac nodal region,
- Obturator nodal region,
- Pudendal nodal region,
- Inguinofemoral nodal region,
- Presacral nodal region,
- Sacral foramina.

According to protocol common iliac, internal iliac, external iliac, obturator and presacral nodal regions should be outlined. Pudendal and inguinofemoral regions should not be included. There was no guidance regarding sacral foramina inclusion. This therefore represents each centres' local practice.

The following were reviewed

- Most superior CTV1 + 2 extent, representing aortic bifurcation,
- Most inferior CTV1 + 2 extent, representing length of vagina included,
- Overlap with muscle/bone,
- Spaces laterally between CTV and muscle/bone.

Neither test case had vaginal tumour extension. The upper half of vagina should therefore be outlined. Muscle and bone should be edited out of CTV with no gaps laterally between pelvic sidewall muscle and/or bone and CTV.

Statistical analysis

Mean, standard deviations and 95% confidence intervals were calculated following Q–Q plots review confirming normality using IBM SPSS Statistics 22. One sample *t*-tests were calculated versus GSCTV1 + 2 to assess for variation between centres. Percentage of centres including specified anatomical regions was calculated and 95% confidence intervals were derived using Exact Confidence Limits for p tables.

Results

Gold standard validation

The STAPLE algorithm created a larger CTV than the TMG for both cases. Case 1 STAPLE volume was 647 cc versus 598 cc (TMG). Case 2 STAPLE volume was 773 cc versus 735 cc (TMG). Mean volume of all centres' CTVs was lower at 518 cc for case 1 (95% confidence interval (CI) 483 cc–553 cc) and 629 cc for case 2 (95% CI 592 cc–666 cc). The TMG and STAPLE volumes are not within these 95% CIs; the TMG outline was 45 cc (case 1) and 69 cc (case 2) larger than the 95% CI upper limit.

The superior border of the TMG and STAPLE outlines for both cases were within 0.25 cm of each other. The extreme points along each axis were within 0.3 cm (case 1) and 0.5 cm (case 2), suggesting similarity. The JCI was 0.76 (case 1) and 0.79 (case 2), corresponding to approximately 86% and 89% overlap. The only discrepancies between areas included in CTV were sacral foramina and pudendal nodal region. The most evident variation between the TMG and STAPLE outline was the length of vagina included. The TMG outline included a longer proportion of vagina; 0.75 cm

Please cite this article in press as: Eminowicz G, McCormack M. Variability of clinical target volume delineation for definitive radiotherapy in cervix cancer. Radiother Oncol (2015), http://dx.doi.org/10.1016/j.radonc.2015.10.007