

RESEARCH

Open Access



Considerable interobserver variation in delineation of pancreatic cancer on 3DCT and 4DCT: a multi-institutional study

Eva Versteijne^{1*†}, Oliver J. Gurney-Champion^{1,2†}, Astrid van der Horst¹, Eelco Lens¹, M. Willemijn Kolff¹, Jeroen Buijsen³, Gati Ebrahimi¹, Karen J. Neelis⁴, Coen R. N. Rasch¹, Jaap Stoker², Marcel van Herk⁵, Arjan Bel¹ and Geertjan van Tienhoven¹

Abstract

Background: The delineation of pancreatic tumors on CT is challenging. In this study, we quantified the interobserver variation for pancreatic tumor delineation on 3DCT as well as on 4DCT.

Methods: Eight observers (radiation oncologists) from six institutions delineated pancreatic tumors of four patients with (borderline) resectable pancreatic cancer. The study consisted of two stages. In the 3DCT-stage, the gross tumor volume (GTV) was delineated on a contrast-enhanced scan. In the 4DCT-stage, the internal GTV (iGTV) was delineated, accounting for the respiratory motion. We calculated the volumes of the (i)GTV, the overlap of the delineated volumes (expressed as generalized conformity index: Cl_{gen}), the local observer variation (local standard deviation: SD) and the overall observer variation (overall SD). We compared these results between GTVs and iGTVs. Additionally, observers were asked to fill out a questionnaire concerning the difficulty of the delineation and their experience in delineating pancreatic tumors.

Results: The ratios of the largest to the smallest delineated GTV and iGTV within the same patient were 6.8 and 16.5, respectively. As the iGTV incorporates the GTV during all respiratory phases, the mean volumes of the iGTV (40.07 cm^3) were larger than those of the GTV (29.91 cm^3). For all patients, Cl_{gen} was larger for the iGTV than for the GTV. The mean overall observer variation (root-mean-square of all local SDs over four patients) was 0.63 cm and 0.80 cm for GTV and iGTV, respectively. The largest local observer variations were seen close to biliary stents and suspicious pathological enlarged lymph nodes, as some observers included them and some did not. This variation was more pronounced for the iGTV than for the GTV. The observers rated the 3DCT-stage and 4DCT-stage equally difficult and treated on average three to four pancreatic cancer patients per year.

Conclusions: A considerable interobserver variation in delineation of pancreatic tumors was observed. This variation was larger for 4D than for 3D delineation. The largest local observer variation was found around biliary stents and suspicious pathological enlarged lymph nodes.

Keywords: Pancreatic cancer, Delineation, Interobserver variation

* Correspondence: e.versteijne@amc.uva.nl

†Equal contributors

¹Department of Radiation Oncology, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Full list of author information is available at the end of the article



Background

The aim of radiotherapy is delivering a high radiation dose to the tumor while minimizing the dose to organs at risk (OARs). For pancreatic tumors, this is challenging due to day-to-day position variation, respiratory motion, and uncertainties in delineation of the tumor [1–4].

The radiation oncologist can delineate the gross tumor volume (GTV) on a three-dimensional CT (3DCT). The GTV is expanded with a margin to account for microscopic extensions, resulting in the clinical target volume (CTV). For the remaining uncertainties, such as organ motion and set-up uncertainties, an additional margin is added to form the planning target volume (PTV). Nowadays, a four-dimensional CT (4DCT) scan is increasingly used to account for tumor motion during respiration [1, 2], for example combined with the internal target volume (ITV) [5] or mid-ventilation approach [6]. For pancreatic cancer patients treated at our department, we combine 4DCT with a modified ITV approach. In this approach, the radiation oncologist delineates the GTV on the average scan of the 4DCT and expands that on all respiratory phases of the 4DCT to generate an internal GTV (iGTV). A 5 mm margin is then added to define the internal CTV (iCTV). An additional PTV margin is added, to account for remaining set-up uncertainties. This PTV margin can be smaller compared with 3DCT delineation since respiratory motion uncertainty is accounted for in this 4D approach. In both the 3DCT and 4DCT approaches it is important that appropriate margin size is used as too small a margin leads to under-treatment of the target volume whereas too large a margin leads to unnecessarily high doses to the OARs. The CTV or iCTV to PTV margins currently used to account for the delineation uncertainties in pancreatic cancer are largely based on assumptions of these uncertainties. To investigate whether these assumptions are correct we performed a delineation study.

Previous delineation studies quantified the interobserver delineation uncertainties for several tumor sites [7–11]. These studies resulted in standardized delineation protocols for those organs. For pancreatic cancer, such a protocol is available in both postoperative setting and preoperative setting [12, 13]. In the study of Carvatta et al., standard criteria for CTV delineation of high risk elective lymph node areas in preoperative or definitive treatment with radiotherapy were developed [13]. Both guidelines were promoted and evaluated in a multi-center dummy-run, and showed an acceptable interobserver variation in delineation of these elective CTVs [14]. Only two other multi-institutional studies on the delineation of pancreatic tumors are available [15, 16]. All three studies show large interobserver variation in GTV delineation, with ratios of the largest to the smallest GTV volume of 6.8 [14], 9 [15] and 3 [16]. Two of these studies were quality control studies of a clinical

trial [15, 16]. Those studies only used 3DCT and included 1–2 patients with locally advanced pancreatic cancer [15, 16]. The third was a delineation study which included two patients and only investigated the interobserver variation using 3DCT [14]. All of these studies reported limited quantitative information (i.e., standard deviations, SD and generalized conformity index, CI_{gen}) [14–16].

The aim of this study was to quantify the interobserver variation for GTV (3DCT) and iGTV (4DCT) delineations. The study included four patients with (borderline) resectable pancreatic cancer, and eight radiation oncologists from six institutions.

Methods

Radiation oncologists (observers) from all nine institutions participating in the PREOPANC trial were asked to participate in this delineation study. Eight observers from six institutions actually participated.

Patients' characteristics

The data of four patients with histologically proven (borderline) resectable pancreatic tumors were used and anonymized. All patients gave written informed consent for both the PREOPANC trial (EudraCT number 2012-003181-40) and MIPA (NCT01989000) study and were the first four patients that randomized for preoperative radiochemotherapy at the Academic Medical Center (AMC) within the PREOPANC trial [17]. Both studies were approved by the local medical ethics committee (PREOPANC: Erasmus Medical Center, Rotterdam; MIPA: AMC, Amsterdam) [17]. Preoperative radiochemotherapy consisted of 15 fractions of 2.4 Gy combined with gemcitabine 1000 mg/m² once a week for three weeks, preceded and followed by a modified course of gemcitabine 1000 mg/m², once a week for two weeks. Between the three cycles there was one week rest [17].

CT scans

Diagnostic CT scan

All patients had a contrast-enhanced diagnostic CT scan in the referring hospital, which was considered to be of adequate diagnostic quality by abdominal radiologists from the AMC with extensive experience in pancreatic cancer. The scans included an axial scan in arterial contrast phase (on average 35 s after injection, all patients), venous contrast phase (on average 60 s after injection, patients 1,2 and 4) and/or a portal contrast phase (on average 240 s after injection, patients 1 and 4) with or without reconstructed coronal views. Two experienced radiologists from the AMC reported the studies. The radiology report of patient 2 described two suspicious loco

regional lymph nodes; the radiology report of patient 4 described some enlarged lymph nodes, which were not further characterized.

Planning CT scan

The planning CT scans were obtained at the radiation oncology department of the AMC with a GE LightSpeed RT 16 scanner (General Electric Company, Waukesha, WI) using a standard acquisition protocol (slice thickness of 2.5 mm). Patients were scanned in treatment position: supine on a flat table top with arms raised above their heads.

First, a 3DCT scan was obtained during free breathing after intravenous Iodine contrast injection. During the same CT session, a few minutes after the 3DCT scan, a 4DCT scan was obtained. The patient's breathing motion was monitored and synchronized to the CT acquisition by the respiratory gating system RPM (Real-Time Position Management, Varian Oncology Systems, Palo Alto, CA). For the 4DCT, images were captured during continuous respiration and divided into ten respiratory bins, resulting in ten image sets of the respiratory cycle. Also, a maximum intensity projection (MIP) and an average intensity projection (Ave-IP) were reconstructed from the ten phase scans. The planning CT scan was obtained during the first modified course of gemcitabine (mean eight days after the first administration of gemcitabine), and on average six weeks (46–62 days, mean 53 days) after the diagnostic CT. The 3DCT and 4DCT scans were registered to each other but not to the diagnostic CT scans.

Fiducial markers and biliary stents

All four patients had intratumoral fiducial markers, which were placed under the guidance of endoscopic ultrasound (EUS), for position verification during radiotherapy [18, 19]. Patients 1, 3 and 4 had a pancreatic head tumor and had received three intratumoral Visicoil fiducial markers (RadioMed, Barlett, TN). For patient 2, two Gold Anchor fiducial markers (Naslund Medical AB, Huddinge, Sweden) and one Visicoil fiducial marker had been placed, but mistakenly in the pancreas head instead of in the corpus tumor. Also, all patients had biliary drainage: patients 1–3 had fully covered metal biliary stents, patient 4 had external percutaneous biliary drainage. All markers, biliary stents and percutaneous biliary drainage had been placed after the diagnostic CT scans and were thus only visible on the planning CT scan.

Delineation software

The Big Brother software, dedicated to radiotherapy delineation studies was used, recording delineations as well as observer-computer interactions [8]. Each observer

received a USB stick containing all CT scans, the radiology report, the Big Brother software, and delineation instructions. These instructions were identical to those in the PREOPANC trial protocol [17].

Delineation protocol

The study consisted of a 3DCT-stage and a 4DCT-stage.

In the 3DCT-stage, the observers were asked to delineate the GTV on the 3DCT scan, which was displayed on the main window. The GTV was defined as the macroscopically visible tumor and neighboring suspicious pathological lymph nodes. A separate window was available for viewing the diagnostic CT scans. A margin of 5 mm was automatically applied to create the CTV.

In the 4DCT-stage, the Ave-IP reconstruction was displayed in the main window. The observers were asked to delineate the GTV on the Ave-IP reconstruction and then create an iGTV defined as the volume encompassing the GTV on all ten respiratory phase image sets of the 4DCT. The diagnostic CT scan, 3DCT scan, and remaining 4DCT images including the MIP reconstruction were available in a separate window. As the 3DCT and 4DCT scans from the planning CT were obtained in the same session, the 3DCT and 4DCT scans were linked to the Ave-IP reconstruction displayed in the main window. Furthermore, a copy of the cursor (linked cursor) was displayed at the corresponding location in the secondary window when these scans were displayed. Once finished with the iGTV delineations, a margin of 5 mm was automatically applied to create the iCTV. Completed delineations were sent back to the investigators by email.

Questionnaire

Observers were asked to fill out a questionnaire containing eight questions about the delineation process (Additional file 1). These multiple choice questions about the delineation process included answers ranging from very easy to very difficult in five steps. In addition, there were three questions about the experience of the observer in delineating pancreatic tumors as well as the number of pancreatic cancer patients the observers treated yearly within and outside the PREOPANC study (Additional file 1).

Data analysis

The data were analyzed using the Big Brother software [8]. The following analyses were repeated for the GTV, iGTV, CTV and iCTV data.

Scatterplots were generated in GraphPad Prism (version 5.00, GraphPad Software, San Diego, CA) to present the range of delineated volumes. Using the Big Brother software we calculated the average volume of the (i)GTV and

CI_{gen} for each patient [20]. The CI_{gen} is a measure of overlap of the delineated volumes and is defined as the ratio of the sum over all observer pairs of the volumes common to both observers and the sum over all observer pairs of the encompassing volumes (volume delineated by at least one of the two observers) [20]. CI_{gen} ranges from 0–1, where 1 indicates full overlap of the delineated volumes from all observers and 0 indicates no overlap. To assess the accuracy of CI_{gen} we repeated its calculation a number of times equal to the number of observers, leaving out one different observer at each repetition. The range of results from this leave-one-out procedure was reported. To test for significant differences in average volumes, we used a two-sided Wilcoxon signed-rank test (32 pairs, significance level $\alpha = 0.05$) using SPSS (version 22.0.0.2, IBM, New York).

To determine the local observer delineation variation per specific area of the (i)GTV/(i)CTV, we calculated for each patient the median surface, i.e., the surface of the volume that was included by at least 50% of the observers [21]. The median surface was sampled with approximately equidistant points at ~ 0.5 mm distance. For each point on the median surface, the perpendicular distance to each delineated (i)GTV/(i)CTV was measured. When a delineated surface was not within 2 cm perpendicular to a point on the median surface, the closest distance from that delineated surface to the reference point on the median surface was used instead. For each point on the median surface, the local observer variation was calculated, defined as the SD of the obtained distances at that point (local SD). Per patient, the overall observer variation (overall SD) was calculated. The overall SD was defined as the root-mean-square of the local SDs. Similar as for the CI_{gen} , the overall SD was repetitively calculated in a leave-one-out procedure and the range was reported.

The answers to the questionnaire were plotted in a scatterplot using GraphPad Prism and the ratings of the difficulty of the delineation between both stages were compared.

Results

Eight observers from six different institutions submitted all GTV and iGTV delineations. The analyses of the delineations reported in this results section were performed on the (i)GTV. The results from the (i)CTV are presented in Additional file 2.

Delineations

Visual inspection of the delineations revealed considerable interobserver variations (Figs. 1 and 2). The ratio of the largest to the smallest delineated GTV and iGTV was 6.8 and 16.5, respectively, both in patient 3. The iGTV volumes were significantly larger than the GTV volumes by 34% ($P = 0.036$). However, for two observers, the

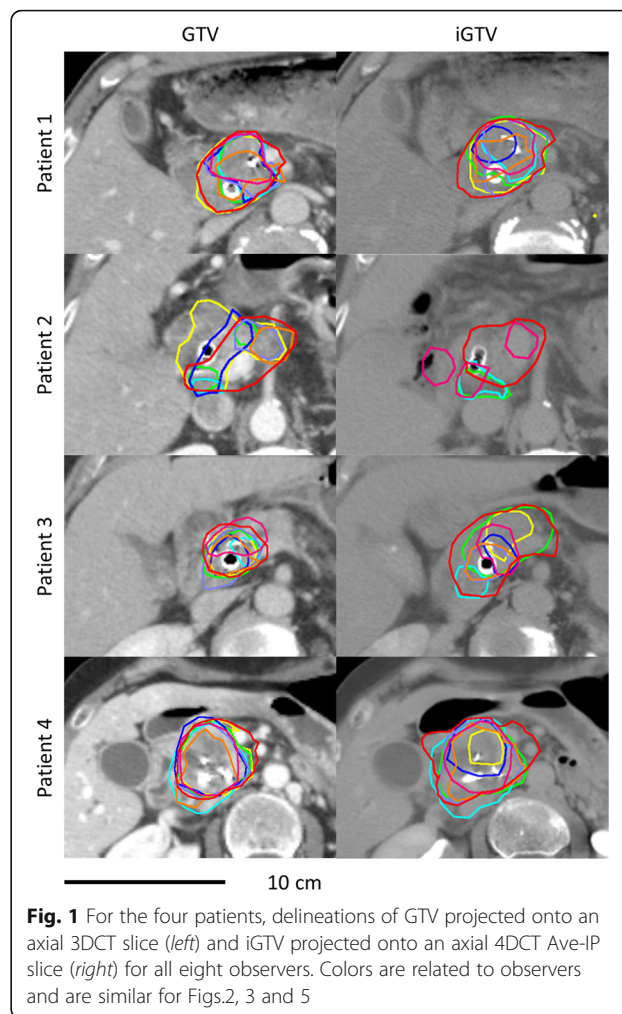


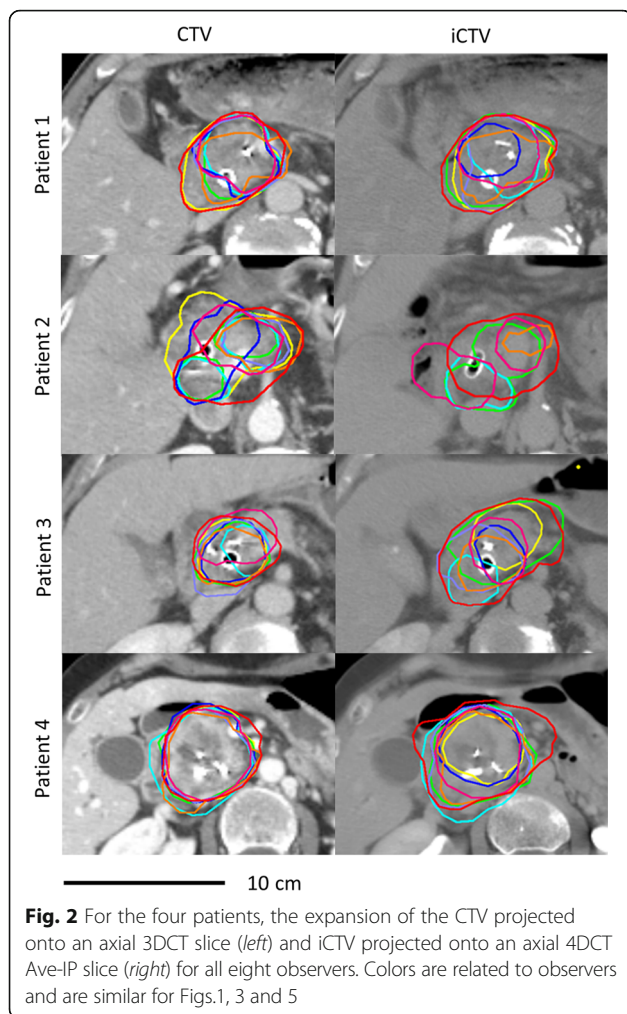
Fig. 1 For the four patients, delineations of GTV projected onto an axial 3DCT slice (left) and iGTV projected onto an axial 4DCT Ave-IP slice (right) for all eight observers. Colors are related to observers and are similar for Figs. 2, 3 and 5

delineated iGTV was smaller than the delineated GTV in all four patients (observers 2 and 5; Fig. 3) and for patient 2 two additional observers (6 and 7) also delineated a smaller iGTV than GTV. Observer 7 reported that his/her iGTV was not based on the ten separate respiratory phases of the 4DCT, due to poor image quality. As the iGTV was delineated on the Ave-IP of the 4DCT, the iGTV still contained 4DCT information. The CI_{gen} was larger for the GTV (mean $CI_{gen} = 0.37$) than for the iGTV (mean $CI_{gen} = 0.27$) for all four patients, indicating a better overlap of volumes in 3D delineation than in 4D delineation (Table 1).

Local observer variation

The local observer variation (local SD) reflects the variation locally projected on the (i)GTV.

There was a large local SD at the laterodorsal borders of the GTV and iGTV of patients 1–3 (Figs. 1 and 4a–b), reflecting the location of the biliary stent. Some observers did, and some did not include the biliary stent in the GTV/iGTV. The biliary stent was included most



often in patient 3: by six observers in the GTV and by four observers in the iGTV (Additional file 3). Especially in patient 2, large local variation was seen. The suspicious pathologically enlarged lymph node in the portocaval space was incorporated in the GTV by five (observers 1-5) and in the iGTV by four (observers 1,3,4 and 8) observers. The suspicious pathologically enlarged lymph node along

the common hepatic artery was included in the GTV by three (observers 2–4) and in the iGTV by two (observers 3 and 4) observers (Additional file 3). Also for patient 2, only observer 2 included all the misplaced fiducial markers in the GTV and only observer 1 included all the fiducial markers in the iGTV. For all patients, there was some variation in including the fiducial markers in the delineated volume (Additional file 3). Finally, the caudal side of tumors had larger local SDs than the other areas of the tumors (Fig. 4a–d).

Overall observer variation

The overall observer variation, represented by the overall SDs of the (i)GTV, was smaller for the GTV delineations (SD = 0.63 cm) compared with the iGTV delineations (SD = 0.80 cm) for all four patients (Table 1). Due to the observed discrepancy in including the suspicious pathologically enlarged lymph nodes in patient 2, we recalculated the overall SD while excluding the portocaval lymph node; the overall SD decreased from 0.84 to 0.72 cm for the GTV and from 0.90 to 0.49 cm for the iGTV.

Questionnaire

Seven observers filled out the structured part of the questionnaire; eight observers the open questions. With a mean score of 3.6 for the difficulty of the delineations in both the 3DCT-stage and 4DCT-stage (Fig. 5), the observers did not consider the iGTV (4DCT) delineation more difficult than the GTV (3DCT) delineation. Of the eight observers that filled out the open questions, one radiation oncologist only just started to treat patients with pancreatic cancer. The remaining seven observers treat on average three to four pancreatic cancer patients per year at their institution (range 1–7.5) and on average they had 5.4 years of experience in delineating pancreatic tumors (range 2–12.5). On average, the observers treated one patient (range 0–4) with pancreatic cancer within the PREOPANC trial. Two observers mentioned in the remarks section that the long interval between the diagnostic scan and planning CT scan (average six weeks) made interpretation more challenging.

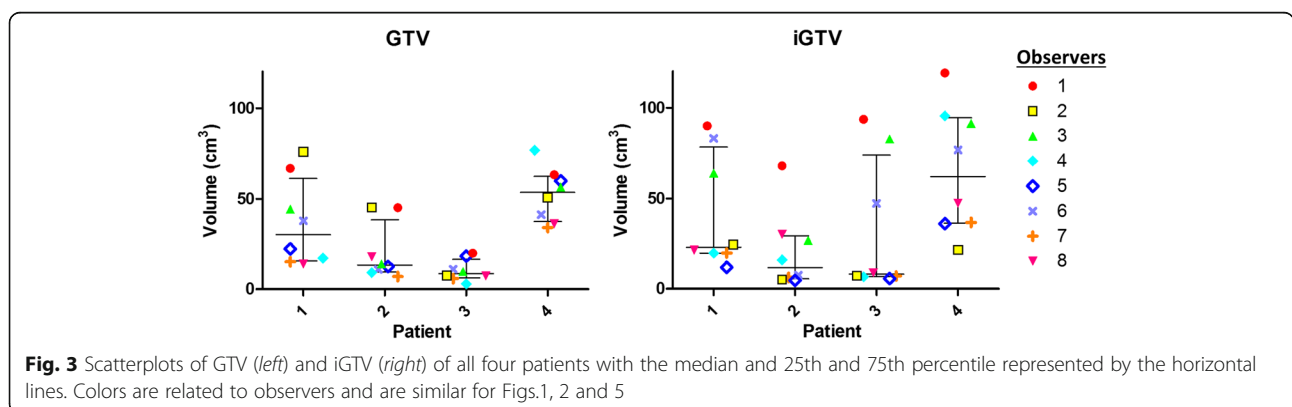


Table 1 The average delineated target volumes, overall SDs and CI_{gen} for all 4 patients

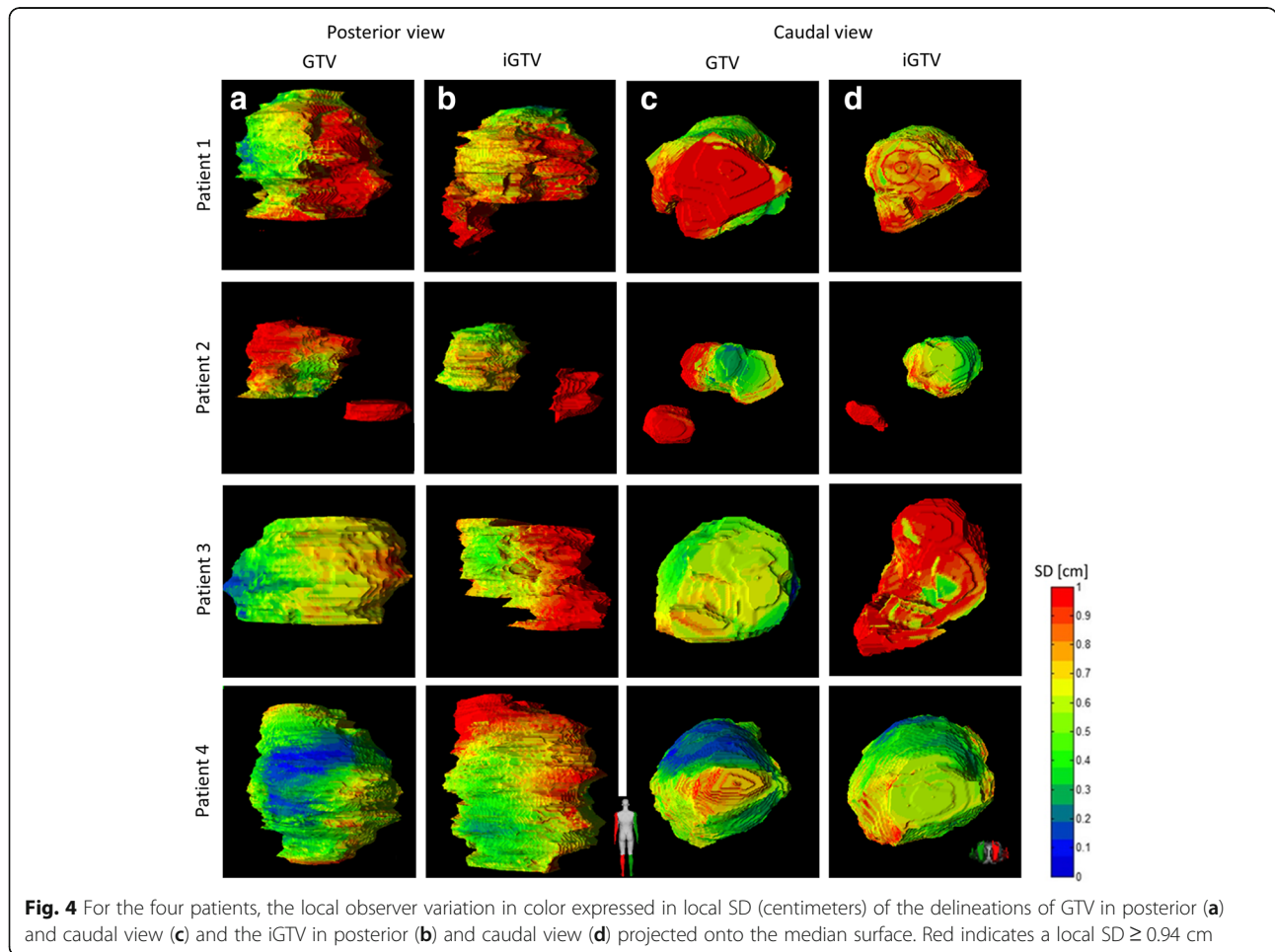
Patient		GTV (range ^a)	iGTV (range ^a)
1	Average volume (cm ³)	36.71 (14.02–75.87)	41.80 (11.85–89.99)
	Overall SD (cm)	0.70 (0.47–0.72)	0.71 (0.60–0.72)
	CI_{gen}	0.34 (0.31–0.37)	0.29 (0.26–0.31)
2	Average volume (cm ³)	20.26 (7.06–45.21)	20.57 (4.67–67.86)
	Overall SD (cm)	0.84 (0.70–0.88)	0.90 (0.37–0.90)
	CI_{gen}	0.22 (0.20–0.27)	0.20 (0.17–0.27)
3	Average volume (cm ³)	10.36 (2.91–19.92)	32.38 (5.67–93.58)
	Overall SD (cm)	0.48 (0.42–0.51)	0.89 (0.77–0.94)
	CI_{gen}	0.34 (0.30–0.37)	0.16 (0.12–0.19)
4	Average volume (cm ³)	52.32 (34.18–76.72)	65.52 (21.48–119.09)
	Overall SD (cm)	0.43 (0.38–0.44)	0.68 (0.58–0.70)
	CI_{gen}	0.59 (0.57–0.62)	0.45 (0.42–0.50)
Overall for all patients	Average volume (cm ³)	29.91	40.07 ^c ($P = 0.036$)
	Overall SD (cm) ^b	0.63	0.80
	CI_{gen}	0.37	0.27

Abbreviations: *GTV* gross tumor volume, *iGTV* internal gross tumor volume, *SD* standard deviation, CI_{gen} generalized conformity index

^aRange over eight delineations (average volume) or over results of the leave-one-out analysis (overall SD and CI_{gen})

^bNote that this overall SD was calculated as the root-mean-square of the four overall SDs from the four patients

^cTwo-sided Wilcoxon signed-rank test



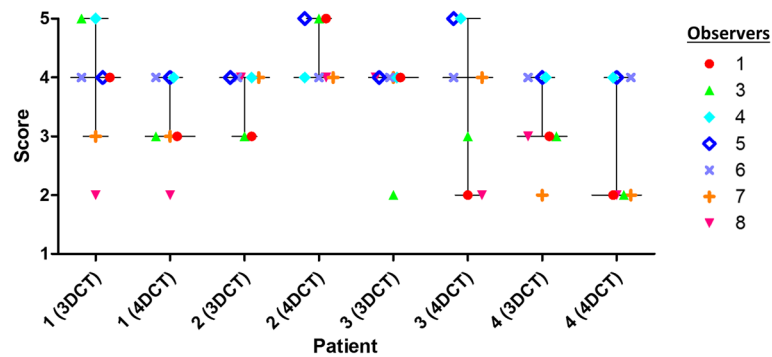


Fig. 5 Scatterplot of the delineation difficulty rating by seven observers (observer 2 missing), showing the range, the median, 25th, 75th percentile for eight delineations. Rating varied between 1 = very easy; 2 = easy; 3 = moderate; 4 = difficult; 5 = very difficult. Colors are related to observers and are similar for Figs. 1, 2 and 3

Discussion

This multi-institutional delineation study is the first to quantify the interobserver variation on both 3DCT and 4DCT. Also, contrary to earlier studies, this study is performed with more than two patients with (borderline) resectable pancreatic cancer. A considerable variation among observers was observed in both GTV (3DCT) and iGTV (4DCT) delineations. The ratio of the largest to the smallest delineated volume was far larger for iGTV than for GTV, with significantly larger average volumes for the iGTV. Furthermore, the GTV delineations had larger CI_{gen} and smaller overall SDs in all patients compared to the iGTV. The largest variation in delineation was seen close to biliary stents and suspicious pathologically enlarged lymph nodes. Previous studies in pancreatic cancer also showed a large interobserver variation on 3DCT with a comparable ratio of largest to smallest GTV of 3–9 [14–16]. The observed interobserver variation is large compared to studies performed in several other organs such as breast, larynx, and rectal cancer, which reported a CI of 0.6–0.82 [9, 22, 23].

The average iGTV volumes were significantly larger than the GTV volumes by 34%. This is similar compared to previous studies in pancreatic cancer, where the iGTV was 25–27.6% larger than the GTV [2, 24]. As the iGTV should incorporate the GTV in all respiratory phases, this result can be expected. However, unexpectedly, in several cases in our study, observers delineated a smaller iGTV than GTV. This may be a result of a large intraobserver variation, which was not specifically assessed in this study. Alternatively, it could be a result of a difference in image quality between the 3DCT and 4DCT images. It is known that inaccuracies in delineation of the tumor may be due to poorly defined tumor edges on the planning CT images [1, 2, 14–16].

The 4DCT delineations had a larger interobserver variation than the 3D delineations, as reflected in the

larger overall SDs of the iGTV compared to the GTV in all four patients and the smaller CI_{gen} . This may be the result of poor visibility of the tumor on the various respiratory phases of the 4DCT. Poor visibility can lead to bigger uncertainty and thus larger target volumes and variation in the delineation. The larger interobserver variation on 4DCT counteract the advantage of the ITV concept: accounting for the respiratory motion. Other delineation approaches with improved contrast between tumor and surrounding tissue to define the tumor borders and including the respiration motion should be investigated such as midventilation and particularly the midposition approach. Previous research showed that a midventilation approach results in significant PTV reduction and significant dose reductions to OARs compared to the iGTV approach, although the delineation process had not been investigated yet [6].

The largest local variation was seen at the laterodorsal side of the (i)GTV, corresponding to the location of the biliary stent. Some observers included the stent in the (i)GTV, whereas others excluded the stent. Also, some observers included the stent only in the GTV but not in the iGTV. In the literature, there is no guideline prescribing to include or exclude the biliary stent in the (i)GTV and none was given in the protocol instructions of the PREOPANC trial. The caudal side of the (i)GTV also showed large variations in delineations, similar to a previous study of Caravatta et al. [14].

Also, large variations in the delineation of the suspicious pathologically enlarged lymph nodes around the tumor were seen. The protocol prescribes to include all neighboring suspicious pathological lymph nodes. The reason for the large local variation that was found around suspicious pathologically enlarged lymph nodes could be due to misinterpretation or ambiguity of protocol instructions, or poor compliance with the protocol instructions. This could also result in the wide range of the separate delineations, with ratios of the largest to the smallest GTV and

iGTV of 6.8 and 16.5. For the GTV these ratios are similar compared to previous studies [14–16], for the iGTV there is no data to compare. To increase interobserver agreement, consensus on the delineation of pancreatic tumors, pathologically enlarged lymph nodes, and biliary stents should be achieved among radiation oncologists. The proposed guidelines of the high risk nodal areas and CTV delineation described by Carvatta et al. might be used to reduce delineation variation in elective CTVs [13, 14]. The lack of guidelines concerning the GTV margin could be an important cause of increased variation of the boost CTV compared to the elective CTV [14]. Also in the post-operative setting, guidelines serve to develop appropriate radiation fields in the setting of very difficult anatomy in the postoperative setting and to ensure that areas at risk are included in the field while organs at risk are spared [12]. Previous research in other organs showed that national consensus guidelines and a delineation atlas may result in reduction of the interobserver delineation variation [10, 25]. Especially for a clinical trial, improvement of interobserver agreement is important. The study of Abrams et al. showed that failure to adhere to specified radiation guidelines was associated with inferior survival [26].

To optimize tumor visibility, the repetition of the diagnostic scan in treatment position after stenting and placement of the fiducial markers may be a step forward. For the patients in our study, registration between the diagnostic CT and the planning CT was not performed because of a different position of the patient and a different anatomy as a result of the placement of the biliary stents and fiducial markers between both scans. Image registration between a diagnostic scan and planning CT scan may improve accuracy in target delineation and reduce interobserver variation as seen for other tumor sites [27–29].

It is well known that pancreatic tumors are difficult to distinguish from normal pancreas tissue on diagnostic CT scans [2, 30, 31]. Therefore, exploitation of other imaging modalities, such as MRI and PET-CT may be a step forward to reduce the variation in delineation of pancreatic tumors. Indeed, other studies have shown that additional imaging, such as MRI and PET-CT, may be helpful in the delineation of pancreatic tumors [32, 33].

Limitations

Delineations were only performed once, and we could not investigate the intraobserver variation. Furthermore, we had a limited number of responding observers, and only a limited number of patients were included. Also, the observers had little experience in the delineation of pancreatic tumors, due to the small number of pancreatic cancer patients eligible for radiotherapy. However, this is typical

for many radiation oncologists and hence the found observer variations should be representative for such radiation oncologists.

The time interval between diagnostic CT and planning CT scan was on average six weeks and the patients were not scanned in treatment position; therefore, anatomical changes (including placement of the biliary stent) occurred between both scans and scans were not registered. This made it challenging to delineate the (i)GTV. However, this is a typical situation in clinical practice in many hospitals since the diagnostic CT is obtained before histological diagnosis, while therapeutic measures such as stenting are performed after the diagnostic CT scan. The fiducial markers in patient 2 were mistakenly not placed inside the tumor, which may have put some observers on the wrong track and contributed to the large interobserver variation seen in this patient.

Conclusion

This study showed a considerable interobserver variation in delineation of pancreatic tumors, larger for 4DCT than for 3DCT delineation. The local variation was largest around the biliary stent and suspicious pathologically enlarged lymph nodes. In the future, better guidelines and the addition of other imaging modalities, such as PET or MRI may help decrease observer variation.

Additional files

Additional file 1: Questionnaire. (PDF 258 kb)

Additional file 2: Data of CTV and iCTV. (PDF 464 kb)

Additional file 3: Inclusion of suspicious pathological lymph nodes, stents and fiducials in the delineations. (PDF 279 kb)

Abbreviations

AMC: Academic medical center; Ave-IP: Average intensity projection; Cl_{gen} : Generalized conformity index; CT: Computed tomography; CTV: Clinical target volume; GTV: Gross tumor volume; iCTV: Internal clinical target volume; iGTV: Internal gross tumor volume; ITV: Internal target volume; MIP: Maximum intensity projection; OAR: Organ at risk; PTV: Planning target volume; SD: Standard deviation

Acknowledgements

The authors would like to thank the participating radiation oncologists (Maaïke Berbee, MAASTRO Clinic Maastricht; Gabriel Paardekooper, Isala Clinics Zwolle; Hanne Heerkens, University Medical Center Utrecht; Radboud University Medical Center Nijmegen) who contributed to this delineation study as observers.

Funding

This work was in part supported by the foundation Bergh in het Zadel through the Dutch Cancer Society (KWF Kankerbestrijding) project no. UVA 2011-5271.

Availability of data and materials

All data generated or analyzed during this study are included in this published article (and its additional information files).

Authors' contributions

EV and OJG drafted the manuscript, analyzed and interpreted the data, AvdH and EL made substantial contributions to conception and design of the study, acquisition of data and analysis and were both involved in revising the manuscript critically for important intellectual content. MWK, JB, GE and KN were observers in the study. CRN, JS, AB and GT made substantial contributions to conception and design of the study acquisition of data and analysis; and were involved in revising the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Competing interests

Dr. A. Bel does consultancy work for Elekta and is the project leader of several Elekta-sponsored projects outside of this work. Professor J. Stoker has a research consultancy for MRI in Crohn's disease with Robarts Clinical Trials.

Consent for publication

Not applicable.

Ethics approval and consent to participate

All patients gave written informed consent to participate in both the PREOPANC trial (EudraCT number 2012-003181-40) and MIPA study (NCT01989000). The PREOPANC trial and MIPA study were both approved by the local medical ethics committee (PREOPANC: Erasmus Medical Center, Rotterdam (MEC-2012-249; December 11th 2012) MIPA: Academic Medical Center, Amsterdam (MEC-2013-254; October 28th 2013).

Author details

¹Department of Radiation Oncology, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.

²Department of Radiology, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.

³Department of Radiation Oncology, MAASTRO clinic, Doctor Tanslaan 12, 6229 ET Maastricht, The Netherlands. ⁴Department of Radiation Oncology, Leiden University Medical Center, Leiden University, Postbus 96002300 RC Leiden, The Netherlands. ⁵Faculty of Biology, Medicine & Health, Division of Molecular & Clinical Cancer Sciences, University of Manchester and Christie NHS trust, Oxford Road Manchester, M13 9PL Manchester, United Kingdom.

Received: 15 October 2016 Accepted: 6 February 2017

Published online: 23 March 2017

References

- Mancosu P, Bettinardi V, Passoni P, Gusmini S, Cappio S, Gilardi MC, et al. Contrast enhanced 4D-CT imaging for target volume definition in pancreatic ductal adenocarcinoma. *Radiother Oncol.* 2008;87(3):339–42. doi:10.1016/j.radonc.2008.04.007.
- Cattaneo GM, Passino P, Sangalli G, Slim N, Longobardi B, Mancosu P, et al. Internal target volume defined by contrast-enhanced 4D-CT scan in unresectable pancreatic tumour: evaluation and reproducibility. *Radiother Oncol.* 2010;97(3):525–9. doi:10.1016/j.radonc.2010.08.007.
- Mori S, Hara R, Yanagi T, Sharp GC, Kumagai M, Asakura H, et al. Four-dimensional measurement of intrafractional respiratory motion of pancreatic tumors using a 256 multi-slice CT scanner. *Radiother Oncol.* 2009;92(2):231–7. doi:10.1016/j.radonc.2008.12.015.
- Lens E, van der Horst A, Kroon PS, van Hooft JE, Dávila Fajardo R, Fockens P, et al. Differences in respiratory-induced pancreatic tumor motion between 4D treatment planning CT and daily cone beam CT, measured using intratumoral fiducials. *Acta Oncol.* 2014;53(9):1257–64. doi:10.3109/0284186X.2014.905699.
- ICRU. Prescribing, recording and reporting photon-beam intensity modulated radiation therapy (IMRT) (ICRU Report 83). 2010.
- Lens E, van der Horst A, Versteijne E, van Tienhoven G, Bel A. Dosimetric advantages of midventilation compared with internal target volume for radiation therapy of pancreatic cancer. *Int J Radiat Oncol Biol Phys.* 2015;92(3):675–82.
- van Mourik AM, Elkhuizen PH, 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(3):286–91. doi:10.1016/j.radonc.2010.01.009.
- Steenbakkens RJHM, Duppen JC, Fitton I, Deurloo KE, Zijp 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(2):182–90.
- Genovesi D, Ausili Cefaro 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(2):89–96. doi:10.1016/j.canrad.2013.11.011.
- Nijkamp J, de Haas-Kock DF, 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(1):14–21. doi:10.1016/j.radonc.2011.08.011.
- 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(7):424–30.
- Goodman KA, Regine WF, Dawson LA, Ben-Josef E, Haustermans K, Bosch WR, et al. Consensus panel contouring atlas for the delineation of the clinical target volume in the postoperative treatment of pancreatic cancer. *Int J Radiat Oncol Biol Phys.* 2012;83(3):901–8. doi:10.1016/j.ijrobp.2012.01.022.
- Carvatta L, Sallustio G, Pacelli F, Padula GD, Deodato F, Macchia G, et al. Clinical target volume delineation including elective nodal irradiation in preoperative and definitive radiotherapy of pancreatic cancer. *Radiat Oncol.* 2012;7:86. doi:10.1186/1748-717X-7-86.
- Caravatta L, Macchia G, Mattiucci GC, Sainato A, Cernusco NL, Mantello G, et al. Inter-observer variability of clinical target volume delineation in radiotherapy treatment of pancreatic cancer: a multi-institutional contouring experience. *Radiat Oncol.* 2014;9:198. doi:10.1186/1748-717X-9-198.
- Yamazaki H, Nishiyama K, Tanaka E, Koiwai K, Shikama N, Ito Y, et al. Dummy run for a phase II multi-institute trial of chemoradiotherapy for unresectable pancreatic cancer: inter-observer variance in contour delineation. *Anticancer Res.* 2007;27(4C):2965–71.
- Fokas E, Clifford C, Spezi E, Joseph G, Branagan J, Hurt C, et al. Comparison of investigator-delineated gross tumor volumes and quality assurance in pancreatic cancer: analysis of the pretrial benchmark case for the SCALOP trial. *Radiother Oncol.* 2015;117(3):432–7. doi:10.1016/j.radonc.2015.08.026.
- Versteijne E, van Eijck CH, Punt CJ, Suker M, Zwinderman AH, Dohmen MA, et al. Preoperative radiochemotherapy versus immediate surgery for resectable and borderline resectable pancreatic cancer (PREOPANC trial): study protocol for a multicentre randomized controlled trial. *Trials.* 2016; 17(1):127. doi:10.1186/s13063-016-1262-z.
- Van der Horst A, Wognum S, Dávila Fajardo R, de Jong R, van Hooft JE, Fockens P, et al. Interfractional position variation of pancreatic tumors quantified using intratumoral fiducial markers and daily cone beam computed tomography. *Int J Radiat Oncol Biol Phys.* 2013;87(1):202–8. doi:10.1016/j.ijrobp.2013.05.001.
- Dávila Fajardo R, Lekkerkerker SJ, van der Horst A, Lens E, Bergman JJ, Fockens P, et al. EUS-guided fiducial markers placement with a 22-gauge needle for image-guided radiation therapy in pancreatic cancer. *Gastrointest Endosc.* 2014;79(5):851–5. doi:10.1016/j.gie.2013.12.027.
- Kouwenhoven E, Giezen M, Struikmans H. Measuring the similarity of target volume delineations independent of the number of observers. *Phys Med Biol.* 2009;54(9):2863–73. doi:10.1088/0031-9155/54/9/018.
- Deurloo KE, Steenbakkens RJ, Zijp LJ, de Bois JA, Nowak PJ, Rasch CR, van Herk M. Quantification of shape variation of prostate and seminal vesicles during external beam radiotherapy. *Int J Radiat Oncol Biol Phys.* 2005;61(1):228–38.
- Jager EA, Kasperts N, Caldas-Magalhaes J, Philippens ME, Pameijer FA, Terhaard CH, Raaijmakers CP, et al. GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. *Radiat Oncol.* 2015;10:26. doi:10.1186/s13014-014-0321-4.
- den Hartogh MD, Philippens ME, van Dam IE, Kleynen CE, Tersteeg RJ, Pijnappel RM, et al. MRI and CT imaging for preoperative target volume delineation in breast-conserving therapy. *Radiat Oncol.* 2014;9:63. doi:10.1186/1748-717X-9-63.
- Goldstein SD, Ford EC, Duhon M, McNutt T, Wong J, Herman JM. Use of respiratory-correlated four-dimensional computed tomography to determine acceptable treatment margins for locally advanced pancreatic adenocarcinoma. *Int J Radiat Oncol Biol Phys.* 2010;76(2):597602. doi:10.1016/j.ijrobp.2009.06.009.
- Rasch CR, Steenbakkens RJ, 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. doi:10.1186/1748-717X-5-21.
- Abrams RA, Winter KA, Regine WF, Safran H, Hoffman JP, Lustig R, et al. Failure to adhere to protocol specified radiation therapy guidelines was

- associated with decreased survival in RTOG 9704 - a phase III trial of adjuvant chemotherapy and chemoradiotherapy for patients with resected adenocarcinoma of the pancreas. *Int J Radiat Oncol Biol Phys.* 2012;82(2):809–16. doi:10.1016/j.ijrobp.2010.11.039.
27. Yang DS, Yoon WS, Lee JA, Lee NK, Lee S, Kim CY, et al. The effectiveness of gadolinium MRI to improve target delineation for radiotherapy in hepatocellular carcinoma: a comparative study of rigid image registration techniques. *Phys Med.* 2014;30(6):676–81. doi:10.1016/j.ejmp.2014.05.003.
 28. Djan I, Petrovic B, Erak M, Nikolic I, Lucic S. Radiotherapy treatment planning: benefits of CT-MR image registration and fusion in tumor volume delineation. *Vojnosanit Pregl.* 2013;70(8):735–9.
 29. Bird D, Scarsbrook AF, Sykes J, Ramasamy S, Subesinghe M, Carey B, et al. Multimodality imaging with CT, MR and FDG-PET for radiotherapy target volume delineation in oropharyngeal squamous cell carcinoma. *BMC Cancer.* 2015;15:844. doi:10.1186/s12885-015-1867-8.
 30. Fletcher JG, Wiersema MJ, Farrell MA, Fidler JL, Burgart LJ, Koyama T, et al. Pancreatic malignancy: value of arterial, pancreatic, and hepatic phase imaging with multi-detector row CT. *Radiology.* 2003;229(1):81–90.
 31. McNulty NJ, Francis IR, Platt JF, Cohan RH, Korobkin M, Gebremariam A. Multi-detector row helical CT of the pancreas: effect of contrast-enhanced multiphase imaging on enhancement of the pancreas, peripancreatic vasculature, and pancreatic adenocarcinoma. *Radiology.* 2001;220(1):97–102.
 32. Li XX, Liu NB, Zhu L, Yuan XK, Yang CW, Ren P, et al. Consequences of additional use of contrast-enhanced (18)F-FDG PET/CT in target volume delineation and dose distribution for pancreatic cancer. *Br J Radiol.* 2015; 88(1051):20140590. doi:10.1259/bjr.20140590.
 33. Dalah E, Moraru I, Paulson E, Erickson B, Li XA. Variability of target and normal structure delineation using multimodality imaging for radiation therapy of pancreatic cancer. *Int J Radiat Oncol Biol Phys.* 2014;89(3):633–40. doi:10.1016/j.ijrobp.2014.02.035.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

