MRI of pancreatic cancer for radiotherapy

Gurney-Champion, O.J.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER 2

Delineation on CT

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

Eva Versteijne*
Oliver J Gurney-Champion*
Astrid van der Horst
Eelco Lens
M Willemijn Kolff
Jeroen Buijsen
Gati Ebrahimi
Karen J Neelis
Coen RN Rasch
Jaap Stoker
Marcel van Herk
Arjan Bel
Geertjan van Tienhoven

*both authors contributed equally to this work

submitted
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: $CIGen$), 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 filled 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, $CIGen$ was larger for the iGTV than for the GTV. The mean overall observer variation (root-mean-square of all local $SD$s over four patients) was 0.63 cm and 0.80 cm for the 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.
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 [14].

Radiation oncologist 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 internal 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 the 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 PTV margins currently used to account for the delineation uncertainties in pancreatic cancer are largely based on estimates of these uncertainties. To investigate whether these estimates 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 only available in the postoperative setting [12]. Only three multi-institutional studies on the delineation of pancreatic tumors are available [13-15]. These studies show large interobserver variation in GTV delineation, with ratios of the largest to the smallest GTV volume of 6.8 [13], 9 [14] and 3 [15]. Two of these studies [14, 15] were quality control studies of a clinical trial. Those studies only used 3DCT and included 1–2 patients with locally advanced pancreatic cancer. The third [13], was a delineation study which included two patients and only investigated the interobserver variation using 3DCT. All of these studies [13-15] reported limited quantitative information (i.e. standard deviation, \( SD \) and generalized conformity index, \( CI_{gen} \)).
The aim of this study was to quantify the interobserver variation for GTV (using 3DCT) and iGTV (using 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 study (NCT01989000) and were the first four patients that randomized for preoperative radiochemotherapy at the Academic Medical Center (AMC) within the PREOPANC trial [16]. Both studies were approved by the local medical ethics committee (PREOPANC: Erasmus Medical Center, Rotterdam; MIPA: AMC, Amsterdam) [16]. 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 [16].

CT scans
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 seconds after injection, all patients), venous contrast phase (on average 60 seconds after injection, patients 1,2 and 4) and/or a portal contrast phase (on average 240 seconds after injection, patients 1 and 4) with or without reconstructed coronal views. Two experienced radiologists from the AMC reported the studies. The report of patient 2 described two suspicious loco regional lymph nodes; the report of patient 4 described some (cited in the radiology report) enlarged lymph nodes, which were not further characterized.
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 (on average eight days after the first administration of gemcitabine), and on average six weeks (46–62 days, with a mean of 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 [17, 18]. Patients 1, 3 and 4 had a pancreatic head tumor and had received three intratumoral Visicoil fiducial markers (RadioMed, Bartlett, TN). For patient 2, two Gold Anchor fiducial markers (Naslund Medical AB, Huddinge, Sweden) and one Visicoil fiducial marker had mistakenly been placed 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 recording delineations as well as observer-computer interactions for radiotherapy delineation studies, was used [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 [16].

**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 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 $Cl_{\text{gen}}$ for each patient [19]. The $Cl_{\text{gen}}$ 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) [19]. $Cl_{\text{gen}}$ ranges from 0 to 1, where 1 indicates full overlap of the delineated volumes from all observers and 0 indicates no overlap. To assess the accuracy of $Cl_{\text{gen}}$ we repeated its calculation a number of times equal to the number of observers, leaving out a different observer 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 or (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 [20]. The median surface was sampled with approximately equidistant (0.5 mm) points. For each point on the median surface, the perpendicular distance to each delineated (i)GTV or (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 perpendicular 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 $SD$s. Similar as for the $Cl_{\text{gen}}$, 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. 2.1 and 2.2). 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 delineated iGTV was smaller than the delineated GTV in all four patients (observers 2 and 5; Fig. 2.3) and for patient 2 two additional observers (6 and 7) also delineated a smaller iGTV than GTV. Observer 7 reported that their iGTV was not based on the ten separate respiratory phases of the 4DCT, due to poor image quality. But as the iGTV was delineated on the Ave-IP of the 4DCT, the iGTV still contained 4DCT information. The $C_{\text{gen}}$ was larger for the GTV (mean $C_{\text{gen}} = 0.37$) than for the iGTV (mean $C_{\text{gen}} = 0.27$) for all four patients, indicating a better overlap of volumes in 3D delineation than in 4D delineation (Table 2.1).

Table 2.1. The average delineated volumes, overall SDs and $C_{\text{gen}}$ for all 4 patients.

<table>
<thead>
<tr>
<th>Patient</th>
<th>GTV (range$^*$)</th>
<th>iGTV (range$^*$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average volume, cm$^3$</td>
<td>Overall SD, cm</td>
</tr>
<tr>
<td>1</td>
<td>36.71 (14.02–75.87)</td>
<td>0.70 (0.47–0.72)</td>
</tr>
<tr>
<td>2</td>
<td>20.26 (7.06–45.21)</td>
<td>0.84 (0.70–0.88)</td>
</tr>
<tr>
<td>3</td>
<td>10.36 (2.91–19.92)</td>
<td>0.48 (0.42–0.51)</td>
</tr>
<tr>
<td>4</td>
<td>52.32 (34.18–76.72)</td>
<td>0.43 (0.38–0.44)</td>
</tr>
<tr>
<td>Overall for all patients</td>
<td>29.91</td>
<td>0.63</td>
</tr>
</tbody>
</table>

* Range over eight delineations (average volume) or over results of the leave-one-out analysis (overall SD and $C_{\text{gen}}$).
† Note that this overall SD was calculated as the root-mean-square of the four overall SDs from the four patients.
‡ Two-sided Wilcoxon signed-rank test

Abbreviations: GTV = gross tumor volume; iGTV = internal gross tumor volume; SD = standard deviation; $C_{\text{gen}}$ = generalized conformity index.
Local observer variation
There was a large local SD (local observer variation) at the laterodorsal borders of the GTV and iGTV of patients 1–3 (Figs. 2.1 and 2.4a-b), reflecting the location of the biliary stent. Some observers did, and some did not include the biliary stent in the (i)GTV. 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). Also, the caudal side of tumors had larger local SDs than the other areas of the tumors (Fig. 2.4).

Overall observer variation
The overall observer variation, represented by the overall SDs of the (i)GTV, was smaller for the GTV delineations (SD = 0.84) compared with the iGTV delineations (SD = 0.90) for all four patients (Table 2.1). 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. 2.5), the observers did not consider the iGTV (using 4DCT) delineation more difficult than the GTV (using 3DCT) delineation. Of the eight observers, 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 PREOUPANC trial. Two observers mentioned in the remarks section that the long
Figure 2.1. 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 2.2, 2.3 and 2.5.
Figure 2.2. For the four patients, the expansion of the CTV projected onto an axial 3DCT slice (left) and iCTV projected onto an axial 4DCT Ave-iP slice (right) for all eight observers. Colors are related to observers and are similar for Figs 2.1, 2.3 and 2.5.
Figure 2.4. For the four patients, the local observer variation in color expressed in local SD (centimeters) of the delineations of the GTV in posterior (a) and caudal view (c) and the iGTV in posterior (b) and caudal view (d) projected onto the median surface. Red indicates a local SD ≥ 0.94 cm.

interval between the diagnostic scan and planning CT scan (average six weeks) made interpretation more challenging.
Figure 2.3. Scatterplots of GTV (left) and iGTV (right) of all four patients with the median and 25th and 75th percentile represented by the horizontal lines. Colors are related to observers and are similar for Figs. 2.1, 2.2 and 2.5.

Figure 2.5. Scatterplot of the delineation difficulty rating by seven observer (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 2.1, 2.2 and 2.3.

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 (using 3DCT) and iGTV (using 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\textsubscript{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 [13-15]. The observed interobserver variation is large compared to studies performed in several other organs such as breast, larynx, and lung, which reported a \( CI_{gen} \) of larger than 0.6 [19, 21, 22].

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, 23]. 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 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 CT images [1, 2, 13-15].

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 may 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. [13].

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 these lymph nodes could be due to misinterpretation or ambiguity of protocol instructions, or poor compliance with the protocol instructions. To increase interobserver agreement, consensus on the delineation of pancreatic tumors, pathologically enlarged lymph nodes, and biliary stents should be achieved among radiation oncologists. Previous research in other organs showed that national consensus guidelines and a delineation atlas may result in reduction of the interobserver delineation variation [10, 24].

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 the diagnostic CT scan and planning CT scan may improve accuracy in target delineation and reduce interobserver variation as seen for other tumor sites [25-27].

It is well known that pancreatic tumors are difficult to distinguish from normal pancreatic tissue on CT scans [2, 28-30]. 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 [31].

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. Finally, 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, the addition of other imaging modalities, such as MRI and PET-CT may help decrease observer variation.
References

1. How did you experience the delineation of the pancreatic tumor of patient 1 on 3DCT?
   a. Very Easy
   b. Easy
   c. Moderate
   d. Difficult
   e. Very difficult

2. How did you experience the delineation of the pancreatic tumor of patient 2 on 3DCT?
   a. Very Easy
   b. Easy
   c. Moderate
   d. Difficult
   e. Very difficult

3. How did you experience the delineation of the pancreatic tumor of patient 3 on 3DCT?
   d. Very Easy
   e. Easy
   f. Moderate
   g. Difficult
   h. Very difficult

4. How did you experience the delineation of the pancreatic tumor of patient 4 on 3DCT?
   a. Very Easy
   b. Easy
   c. Moderate
   d. Difficult
   e. Very difficult
5. How did you experience the delineation of the pancreatic tumor of patient 5 on 4DCT?
   a. Very Easy
   b. Easy
   c. Moderate
   d. Difficult
   e. Very difficult

6. How did you experience the delineation of the pancreatic tumor of patient 6 on 4DCT?
   a. Very Easy
   b. Easy
   c. Moderate
   d. Difficult
   e. Very difficult

7. How did you experience the delineation of the pancreatic tumor of patient 7 on 4DCT?
   a. Very Easy
   b. Easy
   c. Moderate
   d. Difficult
   e. Very difficult

8. How did you experience the delineation of the pancreatic tumor of patient 8 on 4DCT?
   a. Very Easy
   b. Easy
   c. Moderate
   d. Difficult
   e. Very difficult

9. How many years’ experience do you have in delineation pancreatic tumors for radiotherapy?

10. How many patients with pancreatic tumors do you treat per year?

11. Have you treated patients within the PREOPANC study?
## Additional file 2

### Data of CTV and iCTV

**Table 2.A.** The volumes, overall standard deviations and conformity indexes of all 4 patients.

<table>
<thead>
<tr>
<th>Patient</th>
<th>CTV (range*)</th>
<th>iCTV (range*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Average volume (cm³)</td>
<td>79.21 (36.04–145.96)</td>
</tr>
<tr>
<td></td>
<td>Overall SD (cm)</td>
<td>0.75 (0.60–0.77)</td>
</tr>
<tr>
<td></td>
<td>CI\textsubscript{gen}</td>
<td>0.43 (0.41–0.46)</td>
</tr>
<tr>
<td>2</td>
<td>Average volume (cm³)</td>
<td>52.99 (24.97–106.90)</td>
</tr>
<tr>
<td></td>
<td>Overall SD (cm)</td>
<td>0.90 (0.80–0.95)</td>
</tr>
<tr>
<td></td>
<td>CI\textsubscript{gen}</td>
<td>0.32 (0.29–0.37)</td>
</tr>
<tr>
<td>3</td>
<td>Average volume (cm³)</td>
<td>28.66 (11.85–46.76)</td>
</tr>
<tr>
<td></td>
<td>Overall SD (cm)</td>
<td>0.52 (0.45–0.55)</td>
</tr>
<tr>
<td></td>
<td>CI\textsubscript{gen}</td>
<td>0.46 (0.43–0.50)</td>
</tr>
<tr>
<td>4</td>
<td>Average volume (cm³)</td>
<td>103.86 (75.13–147.46)</td>
</tr>
<tr>
<td></td>
<td>Overall SD (cm)</td>
<td>0.44 (0.39–0.46)</td>
</tr>
<tr>
<td></td>
<td>CI\textsubscript{gen}</td>
<td>0.65 (0.63–0.67)</td>
</tr>
<tr>
<td>Overall for all patients</td>
<td>Average volume (cm³)†</td>
<td>66.18</td>
</tr>
<tr>
<td></td>
<td>Overall SD (cm)‡</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>CI\textsubscript{gen}</td>
<td>0.46</td>
</tr>
</tbody>
</table>

* Range over 8 delineation (average volume) or results of leave-one-out analysis (overall SD and CI\textsubscript{gen}).

† Note that the overall SD was calculated as the root-mean-square of the four SDs from the four patients

‡ Two-sided Wilcoxon signed-rank test

Abbreviations: CTV = clinical target volume; iCTV = internal clinical target volume; SD = standard deviation; CI\textsubscript{gen} = generalized conformity index.
Figure 2.A. Scatterplots of CTV (left) and iCTV (right) of all four patients with the median, 25th and 75th percentile represented by the horizontal lines. Colors are related to observers and are similar for Figs. 2.1, 2.2, 2.3 and 2.5 in this paper.
Additional file 3

Suspicious pathological lymph nodes, stents, and fiducials

Table 2.B. Number of observers who included the suspicious pathological lymph nodes in the (i)GTV according the diagnostic CT report.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Location suspicious pathological lymph node</th>
<th>GTV</th>
<th>iGTV</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Portocaval</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Along common hepatic artery</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Along tumor, not characterized</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2.C. Number of observers who included the biliary stent or percutaneous biliary drainage in the (i)GTV at least 50% of the stent in at least three slices.

<table>
<thead>
<tr>
<th>Patient</th>
<th>GTV-CT</th>
<th>iGTV-CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2.D. The number of (i)GTV including the fiducial over the number of fiducials multiplied by number of observers (3 × 8 = 24). Partly included fiducials were counted as delineated.

<table>
<thead>
<tr>
<th>Patient</th>
<th>GTV-CT</th>
<th>iGTV-CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24/24 (100%)</td>
<td>23/24 (100%)</td>
</tr>
<tr>
<td>2*</td>
<td>5/24 (21%)</td>
<td>4/24 (17%)</td>
</tr>
<tr>
<td>3</td>
<td>22/24 (92%)</td>
<td>16/24 (67%)</td>
</tr>
<tr>
<td>4</td>
<td>24/24 (100%)</td>
<td>21/24 (88%)</td>
</tr>
</tbody>
</table>

* In patient 2, the fiducials were mistakenly not implanted in the tumor.