MRI of pancreatic cancer for radiotherapy

Gurney-Champion, O.J.

Citation for published version (APA):
CHAPTER 3

Delineation on CT+MRI

Addition of MRI improves interobserver variation in CT-based pancreatic tumor delineation

Oliver J Gurney-Champion
Eva Versteijne
Astrid van der Horst
Eelco Lens
Heidi Rütten
Hanne D Heerkens
Gabriel MRM Paardekooper
Maaike Berbee
Coen RN Rasch
Jaap Stoker
Marc RW Engelbrecht
Marcel van Herk
Aart J Nederveen
Remy Klaassen
Hanneke WM van Laarhoven
Geertjan van Tienhoven
Arjan Bel

submitted
Abstract

**Purpose:** To assess whether the addition of magnetic resonance imaging (MRI) alongside the planning computed tomography (CT) scan decreases interobserver variation for target volume delineation in pancreatic cancer patients.

**Methods and Materials:** Eight observers (radiation oncologists) from six institutions delineated the gross tumor volume (GTV) on 3DCT, and internal GTV (iGTV) on 4DCT of four pancreatic cancer patients, while MRI was available in a second window (CT+MRI). Variations in volume, generalized conformity index ($CI_{\text{gen}}$), and overall observer variation, expressed as standard deviation ($SD$) of the distances between delineated surfaces, were analyzed. $CI_{\text{gen}}$ is a measure of overlap of the delineated (i)GTVs ($1 = $ full overlap, $0 = $ no overlap). A Wilcoxon signed-rank test (significance level $\alpha = 0.05$) was used to compare these results to those from an earlier study that assessed the interobserver variation by the same observers on the same patients on CT without MRI (CT-only).

**Results:** The maximum ratios between delineated volumes within a patient were 6.1 and 22.4 for the GTV (using 3DCT) and iGTV (using 4DCT), respectively. The average (root-mean-square) overall observer variations were $SD = 0.41$ cm (GTV) and $SD = 0.73$ cm (iGTV). The mean $CI_{\text{gen}}$ was 0.36 for GTV and 0.37 for iGTV. When compared to the (i)GTV delineated on CT-only, the mean volumes of the (i)GTV on CT+MRI were significantly smaller (32%, $p < 0.0005$). The median volumes of the (i)GTV on CT+MRI were included for 97% and 92% in the median volumes of the GTV and iGTV on CT, respectively. Furthermore, CT+MRI showed smaller overall observer variations (GTV and iGTV root-mean-square $SD = 0.59$ cm) in six out of eight delineated structures compared to CT-only (root-mean-square $SD = 0.72$ cm). Finally, although smaller volumes were delineated on CT+MRI, the $CI_{\text{gen}}$ was similar in both studies.

**Conclusion:** The availability of MRI images during target delineation reduced the interobserver variation for 3DCT and 4DCT delineation of target volumes in pancreatic cancer.
Introduction

Radiotherapy for pancreatic cancer aims at delivering a high radiation dose to the tumor while minimizing the dose delivered to the organs at risk (OARs). Several developments have improved the accuracy of radiotherapy of pancreatic cancer. The introduction of intensity modulated radiotherapy and volumetric modulated arc therapy enabled steep dose gradients close to the tumor, reducing the dose to OARs [1, 2]. Also, patient alignment has greatly improved with the introduction of intratumoral fiducial markers combined with daily cone-beam computed tomography (CBCT) [3, 4]. Furthermore, motion management has been improved with the introduction of 4D computed tomography (CT) [5] and breath-holding [6]. One of the remaining major challenges in radiotherapy for pancreatic cancer patients is precise and accurate target volume definition [7-9].

So far, studies on delineation accuracy investigated delineations on 3DCT in 1–2 patients with 11–25 observers. However, no quantitative information was reported on the conformity of the delineated volumes, or on local and overall observer variations expressed as standard deviations (SD). Such parameters are especially relevant for determining appropriate planning treatment volume margins [10] and quantifying improvement in interobserver variation after intervention [11, 12]. In an earlier study with eight observers and four patients, we confirmed a substantial interobserver variation when the tumor was delineated on 3DCT and 4DCT based on these parameters [13].

One of the reasons for this considerable interobserver variation may be the poor contrast between pancreatic tumors and the surrounding tissue on CT images. Potentially, MRI provides better tumor visibility [14-18]. For several other organs, MRI has been shown to improve interobserver variation [19-22]. For pancreatic tumors, delineations based on MRI were only studied in single institute studies (1–3 observers), which did not quantify the interobserver variation [23, 24]. To our knowledge, no multi-center study that assesses the value of MRI for target volume delineation in pancreatic cancer patients is available.

The aim of this exploratory study was to evaluate whether there is an added value in offering MRI alongside the planning CT for delineation of the target volume in pancreatic cancer patients. In this study, we quantify the interobserver variation and compare it to our previously published results [13] on CT-based delineation with the same eight observers and four patients.
Materials and Methods
In our previous study (CT-only) [13], eight observers (radiation oncologists) from six Dutch institutions participating in the PREOPANC trial [25] delineated target volumes of four pancreatic cancer patients using only diagnostic and planning CTs. In the current study (CT+MRI), we asked the same observers to repeat the delineations for the same patients, now offering a diagnostic MRI alongside the CTs. All CT-only data presented in this paper as comparison come from the earlier study [13].

Patients
The same patients were selected as in the CT-only study [13]. In that study, we selected the first four patients with histologically proven (borderline) resectable pancreatic ductal adenocarcinoma who participated in the PREOPANC (EudraCT number 2012-003181-40) trial [25] (radiochemotherapy arm) and MIPA (NCT01989000) study. The patients gave written informed consent to both studies, which were approved by the local medical ethics committees (PREOPANC: Erasmus Medical Center, Rotterdam; MIPA: Academic Medical Center, Amsterdam).

Imaging
Patients underwent a diagnostic CT, MRI and planning CT examination. The diagnostic CT scans (contrast-enhanced; CE) were acquired as part of standard patient care at the referring hospitals. Experienced abdominal radiologists from our institution reviewed these scans and considered them adequate for diagnostic purposes.

After diagnosis, all patients received three markers (intratumoral golden fiducial markers) that were visible on CT, but not on MRI [26], as part of our standard treatment protocol [3, 27]. Furthermore, patients 1–3 received metallic biliary stents, and patient 4 received percutaneous biliary drainage. All were placed after the diagnostic CT, but before the MRI and planning CT.

MRI was performed on an Ingenia 3 T scanner (Philips Healthcare, Best, the Netherlands) as part of the MIPA study. Four MRI scans were obtained using various sequences: T1-weighted spoiled gradient echo (T1W GE), CE T1W GE, T2-weighted turbo spin echo (T2W TSE) and diffusion-weighted imaging (DWI) for which the apparent diffusion coefficient map was displayed [28, 29] (examples in Fig. 3.1 e-h). The diagnostic CT and MRI were not registered to the planning CT.

The planning CT scans were acquired at our radiation oncology department on a GE LightSpeed RT 16 scanner (General Electric Company, Waukesha, WI). Two planning CT scans were acquired, a CE 3DCT and a 4DCT. Several image sets were
reconstructed from the 4DCT scan: the ten respiratory phases, average intensity projection (Ave-IP) and maximum intensity projection (MIP). The diagnostic CT and MRI examinations were performed before the start of radiochemotherapy, whereas the planning CT examination was on average eight days (range 6–10 days) after the patients received their first administration of gemcitabine. Further details of all scans are discussed in the Supplementary Materials.

Delineation

Observers delineated the gross tumor volume (GTV) on the 3DCT (CE) and the internal GTV (iGTV) on the Ave-IP reconstruction of the 4DCT. The GTV was defined as the macroscopically visible tumor and pathological lymph nodes. The iGTV was defined as the GTV delineated on the Ave-IP reconstruction, extended to encompass the GTV on the ten respiratory phases of the 4DCT. The current study (CT+MRI), consisted of a 3DCT+MRI and 4DCT+MRI stage, in which the GTV and iGTV, respectively, were delineated under the guidance of MRI. The results were compared to the results from the 3DCT-only and 4DCT-only stages of the CT-only study [13]. For both studies observers received the same instructions on what to delineate, taken from the PREOPANC trial [25].

In both studies, observers received the Big Brother software [30]. The software showed a primary window, in which the (i)GTV was delineated, and a secondary window that could display selected other available images (Table 3.1). When viewed in the secondary window, the displayed slice from the 3DCT and 4DCT image and cursor position, were linked to the primary window (i.e. same slice and a dot indicating cursor position). Furthermore, observers had access to the radiology reports of the diagnostic CT and, in the CT+MRI study, the radiology reports of the MRI. These reports, from experienced abdominal radiologists, described the tumor extent. For the CT report, the associated pathological lymph nodes were also described: two suspicious locoregional lymph nodes in patient 2; “some” (cited) enlarged lymph nodes in patient 4, which were not characterized further.

First, the software for the CT-only study was sent to all observers. Four weeks after an observer returned their CT-only delineations, that observer received a PowerPoint document. In this document, the visibility of pancreatic tumors on the abovementioned MRI scans was discussed for nine pancreatic cancer patients (different from those included in this study). Observers received the software for the CT+MRI study at least six weeks after returning their CT-only delineations.
Table 3.1. Overview of the available images.

<table>
<thead>
<tr>
<th></th>
<th>3DCT+MRI</th>
<th>3DCT-only*</th>
<th>4DCT+MRI</th>
<th>4DCT-only*</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DCT</td>
<td>X</td>
<td>X</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>4DCT Ave-IP</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>4DCT MIP</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>4DCT ten phases</td>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Diagnostic CTs</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>T1W GE</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>CE T1W GE</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>T2W TSE</td>
<td>+</td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>DWI†</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

* These stages are from the CT-only study [13].
† The apparent diffusion coefficient map of the DWI images was shown.
X = images in main window, used for delineation; + = available in secondary window.
Abbreviations: Ave-IP = average intensity projection; MIP = maximum intensity projection;
T1W GE = T1-weighted gradient echo; CE = contrast enhanced; T2W TSE = T2-weighted turbo spin echo;
DWI = diffusion-weighted imaging.

CT+MRI

The analyses were performed using the Big Brother software [30]. A resident radiation oncologist (EV) visually assessed the individually delineated (i)GTVs. She counted the number of observers who included the stents/drain, markers and pathological lymph nodes in the delineated volume. The average volume of the (i)GTVs, and generalized conformity index ($\text{CI}_{\text{gen}}$) [31] per patient was calculated with the Big Brother software. $\text{CI}_{\text{gen}}$ is the sum over all observer pairs of their common volume (delineated by both observers), divided by the sum over all observer pairs of their encompassing volumes (delineated by at least one observer). $\text{CI}_{\text{gen}}$ is a measure of overlap of the (i)GTVs (1 = full overlap, 0 = no overlap). A median surface was defined using Big Brother. This surface was the 3D closed surface comprising the volume that was included in the (i)GTV by at least 50% of the observers. The median surface was sampled with approximately equidistant (0.5 mm) points in Big Brother. For each point, the perpendicular distances from the median surface to the surfaces of the eight individually delineated (i)GTVs were measured. If the surface of a delineated (i)GTV was not within 2 cm, the distance to the closest point on that surface was used instead. The SD over these eight distances was used as a measure of local observer variation (local SD). The overall observer variation (overall SD) was defined as the root-mean-square of all local SDs. Average overall SDs were calculated as root-mean-square over all patients. To assess
the dependence of the $CI_{gen}$ and overall SD on individual observers, we reported the range of values in a leave-one-out analysis. In this approach, the analyses were repeated eight times, successively leaving out one of the observers.

**Comparison CT+MRI with CT-only**

The results were compared to results from the CT-only study [13]. All statistical analyses were performed using SPSS (version 22.0.0.2, IBM, New York). Plots were made with GraphPad Prism (version 5.00, GraphPad Software, San Diego, CA). To test for differences in mean delineated target volumes, $CI_{gen}$ and mean overall SD we used a two-sided Wilcoxon signed-rank test between the CT-only and CT+MRI studies (significance level $\alpha = 0.05$). Histograms of the local SDs per patient were compared between the CT-only study and CT+MRI study. Finally, we compared the 3D median surfaces from both studies and calculated for each median surface from the CT-only study the percent of the volume surrounded by the median surface of the CT+MRI study and vice versa.

**Results**

Observers had on average 10.6 years (range 3–17 years) of experience in radiotherapy. Observer 4 (12 years of experience in radiotherapy) only just started treating pancreatic cancer. The remaining seven had an average of 5.4 years (range 2–12.5 years) of experience treating pancreatic cancer. These observers saw an average of 3.5 pancreatic cancer patients (range 1–7.5 patients) per year. All of the observers had experience with MRI in radiotherapy, and four had experience with MRI for pancreatic cancer.

**CT+MRI**

The maximum ratio between delineated target volumes within one patient was 6.1 for 3DCT+MRI and 22.4 for 4DCT+MRI (Figs. 3.1 a and c). The mean $CI_{gen}$ was 0.36 for 3DCT+MRI and 0.37 for 4DCT+MRI (Table 3.2). The mean overall observer variation expressed by overall SD was 0.41 cm for 3DCT+MRI and 0.73 cm for 4DCT+MRI (root-mean-square over all patients). Despite the instructions that pathological lymph nodes should be included in the (i)GTV, the different lymph nodes were only delineated by 1–4 (range) out of eight observers (supplemental Table 3.B). Furthermore, the (i)GTVs
showed large variations close to stents/drain, in particular in patient 3 in whom four out of eight observers included the stent for 4DCT+MRI and five out of eight for 4DCT+MRI (supplemental Table 3.C).

Table 3.2. Overview of the average volume, \( CI_{\text{gen}} \), and overall observer variation (overall SD).

<table>
<thead>
<tr>
<th></th>
<th>3DCT+MRI</th>
<th>3DCT-only*</th>
<th>4DCT+MRI</th>
<th>4DCT-only *</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average volume (cm³)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient 1</td>
<td>13.3 (6.1–32.5)</td>
<td>36.7 (14.0–75.9)</td>
<td>19.3 (10.5–52.5)</td>
<td>41.8 (11.9–90.0)</td>
</tr>
<tr>
<td>Patient 2</td>
<td>9.1 (4.5–13.8)</td>
<td>20.3 (7.1–45.2)</td>
<td>18.5 (3.0–51.3)</td>
<td>20.6 (4.7–67.9)</td>
</tr>
<tr>
<td>Patient 3</td>
<td>5.7 (1.9–11.8)</td>
<td>10.4 (2.9–19.9)</td>
<td>11.0 (3.2–26.9)</td>
<td>32.4 (5.7–93.6)</td>
</tr>
<tr>
<td>Patient 4</td>
<td>53.4 (34.8–96.0)</td>
<td>52.3 (34.2–76.7)</td>
<td>59.8 (38.2–76.9)</td>
<td>65.5 (21.5–119.1)</td>
</tr>
<tr>
<td><strong>Mean:</strong></td>
<td>20.4</td>
<td>29.9</td>
<td>27.13</td>
<td>40.07</td>
</tr>
<tr>
<td><strong>CI_{\text{gen}}</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient 1</td>
<td>0.31 (0.29–0.39)</td>
<td>0.34 (0.31–0.37)</td>
<td>0.25 (0.23–0.29)</td>
<td>0.29 (0.26–0.31)</td>
</tr>
<tr>
<td>Patient 2</td>
<td>0.34 (0.32–0.42)</td>
<td>0.22 (0.20–0.27)</td>
<td>0.17 (0.15–0.21)</td>
<td>0.20 (0.17–0.27)</td>
</tr>
<tr>
<td>Patient 3</td>
<td>0.23 (0.20–0.36)</td>
<td>0.34 (0.30–0.37)</td>
<td>0.20 (0.17–0.26)</td>
<td>0.16 (0.12–0.19)</td>
</tr>
<tr>
<td>Patient 4</td>
<td>0.56 (0.53–0.61)</td>
<td>0.59 (0.57–0.62)</td>
<td>0.61 (0.59–0.62)</td>
<td>0.45 (0.42–0.50)</td>
</tr>
<tr>
<td><strong>Mean:</strong></td>
<td>0.36</td>
<td>0.37</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td><strong>Overall SD (cm)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient 1</td>
<td>0.43 (0.36–0.44)</td>
<td>0.70 (0.47–0.72)</td>
<td>0.57 (0.53–0.60)</td>
<td>0.71 (0.60–0.72)</td>
</tr>
<tr>
<td>Patient 2</td>
<td>0.40 (0.28–0.43)</td>
<td>0.84 (0.70–0.88)</td>
<td>1.11 (0.87–1.14)</td>
<td>0.90 (0.37–0.90)</td>
</tr>
<tr>
<td>Patient 3</td>
<td>0.40 (0.35–0.40)</td>
<td>0.48 (0.42–0.51)</td>
<td>0.62 (0.50–0.66)</td>
<td>0.89 (0.77–0.94)</td>
</tr>
<tr>
<td>Patient 4</td>
<td>0.43 (0.37–0.45)</td>
<td>0.43 (0.38–0.44)</td>
<td>0.39 (0.35–0.40)</td>
<td>0.68 (0.58–0.70)</td>
</tr>
<tr>
<td><strong>RMS:</strong></td>
<td>0.41</td>
<td>0.63</td>
<td>0.73</td>
<td>0.80</td>
</tr>
</tbody>
</table>

* These results are from the CT-only study (13).

Abbreviations: SD = standard deviation; \( CI_{\text{gen}} \) = generalized conformity index; RMS = root-mean-square.

Between brackets, range of volumes (Average volume) and range of leave-one-out analyses (overall SD and \( CI_{\text{gen}} \)).

**Comparison CT+MRI with CT-only**

Delineated \((i)\)GTVs were 32% smaller when based on CT+MRI compared to CT-only \((p < 0.0005, Z = -3.826, \text{Fig. } 3.1 \text{ a-d, Fig. } 3.2 \text{ and Table } 3.2)\). There was no significant difference in mean \( CI_{\text{gen}} \) between the CT+MRI study \((CI_{\text{gen}} = 0.34)\) and CT-only study \((CI_{\text{gen}} = 0.32; p = 0.844, Z = -0.280)\). The mean overall observer variation was not significantly smaller in the CT+MRI study (root-mean-square over patients and \((i)\)GTV overall SD = 0.59 cm) than in the CT-only study (0.72 cm; \(p = 0.078, Z = -1.820)\).
Figure 3.1. Example slices of GTV delineated on 3DCT+MRI (a) or 3DCT-only [13] (b) and iGTV delineated on the Ave-IP using 4DCT+MRI (c) or 4DCT-only [13] (d). The MRI images (e-h) show the tumor in the manually selected corresponding slice. The apparent diffusion coefficient map from the DWI acquisition is shown (h). Abbreviations: T1W = T1-weighted; GE = gradient echo; CE = contrast enhanced; T2W = T2-weighted; TSE = turbo spin echo; DWI = diffusion-weighted imaging.
However, they were smaller in six out of eight delineated structures (three out of four GTVs, three out of four iGTVs).

Making MRI available decreased the local observer variation particularly in regions that had a large (> 1 cm) local SD in the CT-only study (Fig. 3.3). The histograms of local SD reflect this effect (supplemental Fig. 3.A). These histograms show shorter tails at high local SD for CT+MRI compared to CT-only, especially for the GTV of patients 1 and 2, and the iGTV of patients 3 and 4.

The median delineated surfaces of the (i)GTV of patient 2 each consisted of two separate volumes (except during the 3DCT+MRI stage), one encompassing the main tumor and one encompassing a suspicious portocaval lymph node (Fig. 3.3). As only 4–5 observers delineated this lymph node (supplemental Table 3.B), the local observer variation was large for that part of the median surface (Fig. 3.3, patient 2: local SD > 1 cm). Excluding this lymph node from the median surface resulted in an overall SD = 0.72 cm (leave-one-out range: 0.50–0.75 cm) in the 3DCT-only stage,
0.86 cm (0.44–0.91 cm) in the 4DCT+MRI stage and 0.49 cm (0.37–0.51 cm) in the 4DCT-only stage.

The median surfaces from the CT+MRI study were on average for 97% and 92% included within the median surfaces of the CT-only study for the GTV (using 3DCT) and iGTV (using 4DCT), respectively (Fig. 3.4). Vice versa, the median surfaces from the CT-only study were for 56% and 64% included within the median surface of the CT+MRI study.

Figure 3.3. The local observer variation (local SD) projected on the median surfaces of the (i)GTV for 3DCT+MRI (a), 3DCT-only (b), 4DCT+MRI (c) and 4DCT-only (d). Colors correlate to the local SD, with red indicating local SD ≥ 9.4 mm. Volumes are viewed posterior of the patient (dummy). Note that for patient 2 the portocaval lymph node was included in (b–d). Figures (b) and (d) are from Versteijne et al [13].
The number of observers who delineated a specific lymph node only varied by a maximum of one observer between the two studies (supplemental Table 3.B). An exception was the portocaval lymph node in patient 2, which was delineated by two observers in the 3DCT+MRI stage, instead of 4–5 observers in all other stages. There was a poor agreement in all stages on whether to include stents/drains in the (i)GTV (supplemental Table 3.C). There was a small difference (< 17% for all patients) between the CT+MRI study and CT-only study in the percentage of markers included in the (i)GTV (supplemental Table 3.D), except in patient 3. In patient 3 the markers were less often included for CT+MRI than for CT-only delineations.

Discussion
We are the first to show the benefits of offering MRI alongside the planning CT for delineation of pancreatic tumors in a multi-center setting. When MRI was available, the precision, represented by the overall SD, improved in six out of eight delineated structures. Furthermore, the volumes of the (i)GTV decreased significantly. The overlap of delineated (i)GTVs, represented by $CI_{gen}$, remained similar, despite the
smaller volumes. These findings suggest an extra value of adding MRI for pancreatic tumor delineation.

In our study, as well as in other studies [23, 24], delineated target volumes in pancreatic cancer patients were smaller when (partially) based on MRI than when based on CT only. Similar findings have been reported for other types of cancer [19-21]. There are two possible causes for this decrease in volume. The tumor volume size could be overestimated on CT due to poor tumor contrast and uncertain tumor boundaries. Furthermore, tumor size could be underestimated on (CT+)MRI. Therefore, appropriate clinical target volume (CTV) margins should be determined.

We assessed the number of interactions done in Big Brother with the various MRI sequences and found that observers focused on the CE T1W GE (results not shown). In one study [32], pancreatic tumor sizes were underestimated when assessed on CE MRI. In a different study [23], it was shown that delineations of pancreatic cancer tumors on MRI were larger when based on DWI than when based on CE T1W GE images. More extensive use of the other MRI sequences may improve delineations [18]. Future protocol instructions for MRI-based tumor delineation should be developed taking into account these issues.

The addition of MRI mainly decreased the local SD in regions that had large local observer variation (local SD > 1cm). The fact that less decrease was seen in other regions of low local observer variation could be a result of the MRI images not being registered to the planning CT images. Translating the MRI information to the CT images may be challenging for the observers. Potentially, registration of the CT and MRI images decreases the overall observer variation further, as shown for brain tumors [33]. As the pancreas and other abdominal organs had deformed between the three image sessions, a deformable registration would be required. The use of deformable image registration for this purpose is not widely validated. Therefore, we chose not to register the images. In the future, MRI and CT scans should be acquired in the treatment position, preferably with markers that are visible on both MRI and CT to guide registration [26].

Clear instructions and consensus on what to delineate decreases interobserver variation, as was shown in various organs [12, 11]. In our study, no specific instructions on whether to include stents/drains into the (i)GTV were given. The large variation on including stents suggests the necessity of such instructions. Our delineation instructions did state that pathological lymph nodes should be included in the (i)GTVs. The large variation concerning the inclusion of lymph nodes in the (i)GTV suggests that observer compliance is also important. As these confusions affected both the CT-only and CT+MRI studies (Tables 3.B-3.D), we believe they did not influence the comparison between the CT-only and CT+MRI delineations.
Several limitations were associated with this paper. Due to the relatively large amount of pancreatic tumor delineations done (four patients, four times) in a relatively short time, observers may have improved their skills between both studies. Furthermore, despite the gap of at least six weeks between both studies, observers may have recognized patients during the CT+MRI study. Both factors could result in smaller observer variation for CT+MRI. Furthermore, the MRI was not obtained as part of the radiotherapy treatment. When MRI is obtained for radiation treatment, different settings or sequences may be preferred with higher resolutions, such as e.g. alternating repetition time balanced steady-state free precession imaging, which could be an alternative to the T2W TSE acquisition, with high resolution in 3D [34].

Similar to findings for other organs [19-22], we found that the overall observer variation decreased when MRI was available during delineation. Consequently, potentially smaller planning treatment volume margins can be used for CT+MRI delineations than for CT-only delineations. However, due to the exploratory nature of this research, the overall SD was based on a small patient group and, therefore, does not necessarily represent typical overall SDs for the pancreatic cancer patient population. Therefore, this study should be repeated using more patients to quantify necessary treatment margins.

In conclusion, in this exploratory study the availability of MRI images to CT during target volume delineation for pancreatic cancer improved (decreased) the overall observer variation for six out of eight structures and resulted in smaller delineated volumes compared to CT-only delineation. Yet, large local observer variations existed close to the biliary stent and/or pathological lymph nodes, indicating issues with instructions and compliance.
References


Supplementary Materials A

Acquisitions

Diagnostic CT, planning CT and MRI examinations were held on separate days. On average, the diagnostic CT examination was 53 days (range 46–62 days) before the planning CT examination whereas the MRI examination was 19 days (range 15–21 days) before the planning CT examination.

The diagnostic CT scans were acquired at the referring hospitals. The scans were considered adequate for diagnostic purposes by the dedicated abdominal radiologists from our institution. The diagnostic CT scans were obtained after vascular contrast (iodine) injection and included axial scans in arterial (35 s after injection, all patients), and venous contrast phase (60 s after injection, patients 1, 2 and 4) or portal contrast phase (240 s after injection, patients 1 and 4). In addition, reconstructed coronal views were available for the arterial (patients 1, 2 and 4) and venous (patients 2 and 3) contrast phases.

The planning CT scans were obtained with patients in treatment position: supine on a flat table top with hands above their head. Two planning CT scans were acquired, a 3DCT and 4DCT. The 3DCT scan was acquired during free breathing after contrast injection in the arterial contrast phase and the corresponding 3DCT images represent a snapshot of the respiratory cycle. The 3DCT was reconstructed with a slice thickness of 2.5 mm and an in-plane resolution of 1.3 × 1.3 mm². For the 4DCT scan, the respiratory motion was monitored while data were acquired continuously during free breathing. The respiratory motion was synchronized to the CT acquisition by the respiratory gating system RPM (Real-Time Position Management, Varian Oncology Systems, Palo Alto, CA). The data were sorted according to the respiratory signal and images were reconstructed of the ten respiratory phases. Also, a maxi-mum intensity projection (MIP) and average intensity projection (Ave-IP) were reconstructed from the images from the ten phases. The images reconstructed from the 4DCT scan had a slice thickness of 2.5 mm and an in-plane resolution of 1.0 × 1.0 mm³.

MRI scans were obtained with a 16-channel phased-array anterior coil and a 10-channel phased-array posterior coil. Four MRI scans were acquired using various sequences: T1W GE, CE T1W GE, T2W TSE and DWI (Table 3.A). The T1W GE scan obtained three echoes and used a Dixon reconstruction to obtain water only images. The T1W GE scan was acquired in a single breath-hold and images had a high isotropic resolution, allowing for a correct representation of the anatomy. However, T1W GE images show poor pancreatic tumor contrast. Therefore, an additional CE T1W GE scan was acquired after gadovist 1.0 (Bayer Healthcare, Leverkusen, Germany) administration (0.1 ml/kg; 5 ml/s, followed by a 15 ml saline flush) in the
parenchymal phase (35 seconds after injection), which improves lesion conspicuity. Due to the design of the MIPA study, another gadovist bolus had also been injected approximately 6 minutes before the CE T1W GE scan. On T2W TSE images the tumor is often poorly visible; however, its location can often be derived by finding obstructions of the pancreatic duct, which is clearly visible on these images. Also, these images can show the presence of peripancreatic inflammation. The apparent diffusion coefficient maps were generated from $b = 0$ and 600 mm$^2$/s DWI scans. DWI, and in particular, the apparent diffusion coefficient maps from DWI, give information about the diffusivity of tissue. In general, pancreas carcinomas show increased cellular density, causing decreased diffusivity [28, 29].

Table 3.A. MRI sequence settings.

<table>
<thead>
<tr>
<th></th>
<th>T1W GE</th>
<th>T2W TSE</th>
<th>DWI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (mm$^2$)</td>
<td>1.7 x 1.7</td>
<td>1.3 x 1.6</td>
<td>3 x 3</td>
</tr>
<tr>
<td>Slice thickness (mm)</td>
<td>1.7</td>
<td>5</td>
<td>3.7</td>
</tr>
<tr>
<td>FOV RL x AP x FH (mm$^3$)</td>
<td>400 x 353 x 95</td>
<td>400 x 369 x 145</td>
<td>432 x 108 x 67</td>
</tr>
<tr>
<td>TR/TE/$\Delta$TE (ms)</td>
<td>4.7/1.26/0.9</td>
<td>779/80/–</td>
<td>$&gt; 2200/45/–$</td>
</tr>
<tr>
<td>Parallel imaging</td>
<td>2 (AP)/1.5 (FH)</td>
<td>2 (AP)</td>
<td>1.3 (AP)</td>
</tr>
<tr>
<td>Fat saturation</td>
<td>Dixon</td>
<td>SPAIR</td>
<td>SPIR and gradient reversal</td>
</tr>
<tr>
<td>Flip angle (°)</td>
<td>9 (25 post contrast)</td>
<td>90 (120 refocussing)</td>
<td>90</td>
</tr>
<tr>
<td>Respiratory compensation</td>
<td>1 breath-hold</td>
<td>3 breath-holds</td>
<td>Respiratory triggering using a navigator</td>
</tr>
</tbody>
</table>

Abbreviations: T2W TSE = T2-weighted turbo spin echo; DWI = diffusion-weighted MRI; FOV = field of view; RL = right–left; AP = anterior–posterior; FH = foot–head; TR = repetition time; TE = echo time; $\Delta$TE = increase in TE; SPAIR = spectral attenuated inversion recovery; SPIR = spectral presaturation with inversion recovery.
## Supplementary Materials B

### Lymph nodes, stents/drain, and markers

#### Table 3.B. Number of observers who delineated the lymph nodes.

<table>
<thead>
<tr>
<th>lymph nodes</th>
<th>3DCT+MRI</th>
<th>3DCT-only*</th>
<th>4DCT+MRI</th>
<th>4DCT-only*</th>
</tr>
</thead>
<tbody>
<tr>
<td>patient 2, portocaval</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>patient 2, along common hepatic duct</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>patient 4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

* These results are from the CT-only study [13].

#### Table 3.C. Number of observers who delineated the stent or drain.

<table>
<thead>
<tr>
<th>Patients</th>
<th>3DCT+MRI</th>
<th>3DCT-only*</th>
<th>4DCT+MRI</th>
<th>4DCT-only*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

Number of observers who delineated at least 50% of the stent/drain in at least three slices are reported

* These results are from the CT-only study [13].

#### Table 3.D. Markers located within the delineated volumes.

<table>
<thead>
<tr>
<th></th>
<th>3DCT+MRI</th>
<th>3DCT-only*</th>
<th>4DCT+MRI</th>
<th>4DCT-only*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24/24</td>
<td>(100%)</td>
<td>24/24</td>
<td>(100%)</td>
</tr>
<tr>
<td>2†</td>
<td>1/24</td>
<td>(4%)</td>
<td>5/24</td>
<td>(21%)</td>
</tr>
<tr>
<td>3</td>
<td>11/24</td>
<td>(46%)</td>
<td>22/24</td>
<td>(92%)</td>
</tr>
<tr>
<td>4</td>
<td>24/24</td>
<td>(100%)</td>
<td>24/24</td>
<td>(100%)</td>
</tr>
</tbody>
</table>

The number of volumes including the marker over the number of markers multiplied by number of observers (3 × 8 = 24). Markers that were included partially in the delineated volume were counted as delineated.

* These results are from the CT-only study [13].

† In patient 2, the markers were mistakenly implanted outside the tumor.
Supplementary Materials C

Local standard deviation

Figure 3.A. Normalized histogram of the local SD of the CT-only study (dotted gray line) and CT+MRI study (solid black line). Graphs are based on the local SDs from on average 32,324 (range 9,728–66,504) points on the median surface and normalized to the total surface of the median surface.