Phase II randomised trial

The TRENDY multi-center randomized trial on hepatocellular carcinoma – Trial QA including automated treatment planning and benchmark-case results

Steven J.M. Habraken a,⇑, Abdul Wahab M. Sharfo a, Jeroen Buijsen b, Wilko F.A.R. Verbakel c, Cornelis J.A. Haasbeek c, Michel C. Öllers b, Henrike Westerveld d, Niek van Wieringen d, Onne Reerink e, Enrica Seravalli e, Pètra M. Braam f, Markus Wendling f, Thomas Lacornerie g, Xavier Mirabel g, Reinilde Weytjens h, Lieselotte Depuydt l, Stephanie Tanadini-Lang i, Oliver Riesterer i, Karin Haustermans j, Tom Depuydt j, Roy S. Dwarkasing k, François E.J.A. Willemssen k, Ben J.M. Heijmen a, Alejandra Méndez Romero a

a Erasmus MC Cancer Institute, Radiotherapy, Rotterdam; b MAASTRO Clinic, Department of Radiation Oncology, Maastricht; c VU University Medical Center, Radiation Oncology, Amsterdam; d Academic Medical Center, Radiotherapy, Amsterdam; e University Medical Center Utrecht, Department of Radiotherapy, Utrecht; f Radboud University Medical Center, Radiation Oncology, Nijmegen, The Netherlands; g Oscar Lambret Comprehensive Cancer Center, Academic Radiation Therapy Department, Lille, France; h University Hospital Antwerp, Iridium Kankernetwerk, Radiotherapy, Antwerp, Belgium; i University Hospital Zurich, Department of Radiation Oncology, Zurich, Switzerland; j University Hospitals Leuven, Department of Radiation Oncology, Belgium; k Erasmus MC, Department of Radiology and Nuclear Medicine, Rotterdam, The Netherlands; l GZA Ziekenhuizen / Iridium Kankernetwerk, Radiotherapy, Wilrijk, Belgium

Article info

Article history:
Received 31 May 2017
Received in revised form 7 September 2017
Accepted 9 September 2017
Available online 16 October 2017

Keywords:
QA in clinical trials
Benchmark-case delineation
Benchmark-case treatment planning
Automated treatment planning for plan QA

Abstract

Background and purpose: The TRENDY trial is an international multi-center phase-II study, randomizing hepatocellular carcinoma (HCC) patients between transarterial chemoembolization (TACE) and stereotactic body radiation therapy (SBRT) with a target dose of 48–54 Gy in six fractions. The radiotherapy quality assurance (QA) program, including prospective plan feedback based on automated treatment planning, is described and results are reported.

Materials and methods: Scans of a single patient were used as a benchmark case. Contours submitted by nine participating centers were compared with reference contours. The subsequent planning round was based on a single set of contours. A total of 20 plans from participating centers, including 12 from the benchmark case, 5 from a clinical pilot and 3 from the first study patients, were compared to automatically generated VMAT plans.

Results: For the submitted liver contours, Dice Similarity Coefficients (DSC) with the reference delineation ranged from 0.925 to 0.954. For the GTV, the DSC varied between 0.721 and 0.876. For the 12 plans on the benchmark case, healthy liver normal-tissue complication probabilities (NTCPs) ranged from 0.2% to 22.2% with little correlation between NCTP and PTV-D95% (R2 < 0.3). Four protocol deviations were detected in the set of 20 treatment plans. Comparison with co-planar autoVMAT QA plans revealed these were due to too high target dose and suboptimal planning. Overall, autoVMAT resulted in an average liver NTCP reduction of 2.2 percent point (range: 16.2 percent point to /C0 1.8 percent point, p = 0.03), and lower doses to the healthy liver (p < 0.01) and gastrointestinal organs at risk (p < 0.001).

Conclusions: Delineation variation resulted in feedback to participating centers. Automated treatment planning can play an important role in clinical trials for prospective plan QA as suboptimal plans were detected.

Quality assurance (QA) is essential to clinical trials involving radiation therapy, as protocol violations may seriously impact trial outcome [1–7]. This holds for all trial aspects, including delineation [8–11] and planning [12–14].

The TRENDY trial (registered as NCT02470533 on clinicaltrials.gov) is an international multi-center clinical trial in which patients with hepatocellular carcinoma (HCC) are randomized between transarterial chemoembolization (TACE) with drug-
eluting beads in the standard arm and stereotactic body radiation therapy (SBRT) in the experimental arm. Radiotherapy is delivered in six fractions with a total target dose of 48–54 Gy. The primary endpoint is time to progression. Secondary endpoints include time to local recurrence, response rate (partial or complete), overall survival, toxicity and quality of life. Patients are accrued from eleven centers in four different countries. Radiotherapy is delivered in ten of these centers, all experienced with liver SBRT; patients from one center are referred to Erasmus MC for radiotherapy.

Due to underlying liver disease (cirrhosis), the patient population in the TRENDY trial is prone to developing hepatic toxicity after radiotherapy, and sparing of the healthy liver is crucial in these patients [15]. As radiotherapy of the liver is technologically challenging, an extensive QA program, comparable to that of other recent radiotherapy trials on primary liver cancer [16,17], has been developed. QA guidelines and recommendations are outlined in a QA protocol, included in the supplementary material. The trial QA is coordinated by a QA team, consisting of two radiation oncologists and two medical physicists from different participating centers. The QA team is supported by two radiologists for advice on target definition. Prior to inclusion of patients, centers are supposed to recently have had an external beam output audit. Also, they need to fill out a trial-specific facility questionnaire explaining in detail technical aspects of the applied procedures for treating HCC patients. In addition, centers have to prepare delineation and treatment planning on a benchmark case [18]. During patient accrual, the QA protocol accommodates patient-specific prospective feedback (prior to treatment) on target definition, organ at risk (OAR) delineation and treatment planning. It defines minor and major protocol deviations. When a major deviation is detected during prospective monitoring by the QA team, the participating center needs to first improve the contouring and/or planning before start of treatment. Retrospectively detected major deviations have to be avoided for future patients from the same center. Minor deviations should be avoided when possible. For each submitted patient, a volumetric-modulated arc therapy (VMAT) plan is generated at Erasmus MC on behalf of the QA team, using fully automated multi-criteria plan generation (autoVMAT) [19,20]. AutoVMAT planning results are provided to the treating center. The aim is to provide prospective feedback for as many patients as possible, but at the very least for the first three patients from each center. Depending on feedback on the first patients, subsequent patients can be reviewed either prospectively or retrospectively (after start of treatment). To make prospective feedback practically feasible, participating centers are invited to submit at least three days prior to the planned start of treatment all images, contours and the treatment plan.

This paper contains quantitative analyses of the contouring and treatment planning benchmark-case results and reports on the first experiences with treatment plan feedback based on autoVMAT.

Materials and methods

Target definition and contouring guidelines

The target lesion(s) should be visible on contrast-enhanced CT imaging and are primarily delineated on the arterial contrast phase of the CT scan. Additional diagnostic imaging (MR/PET) is recommended when available. In regions with poor visibility of tumor edges, generous delineation is required to avoid tumor miss. General recommendations for OAR delineation are outlined in the study protocol. Additional published guidelines were supplied to support delineation [21]. OARs that have to be fully delineated in 3D are: liver, kidneys, stomach, heart and gallbladder. For the required calculation of the liver normal-tissue complication probability (NTCP, see below), a healthy liver structure has to be created, defined as the full liver minus the gross tumor volume (GTV), and minus parts of the liver that are not functional such as regions previously treated with radio-frequency ablation (RFA) (if present). Partial delineations are allowed for spinal cord, duodenum, esophagus, and bowel (in case areas of the small and/or large bowel are located close to the tumor or could be located in high dose gradients).

Planning constraints and objectives

In treatment planning, the goal is to irradiate the full PTV with at least 48 Gy, or, if possible within planning constraints, up to 54 Gy, in six fractions. General recommendations for treatment planning and planning constraints are outlined in the trial protocol and summarized in the QA protocol, where also minor and major deviations are defined. The healthy liver is an important OAR. Several planning constraints are used to guarantee safe liver dose delivery, in particular, NTCP ≤ 5%. An NTCP between 5% and 10% qualifies as a minor deviation. For NTCP calculation, the healthy liver dose-volume-histogram (DVH) is first converted into a DVH for 1.5 Gy/fraction dose delivery in every voxel using $\alpha/\beta = 2.5$ Gy. Then the NTCP is calculated using $TD_{50} = 39.8$ Gy, $m = 0.12$, $n = 0.97$ [15,22]. A software implementation for NTCP calculation has been developed and distributed among the participating radiotherapy centers. In addition, the average dose to the healthy parts of the liver should be less than 22 Gy [22] and, based on [23] $\alpha/\beta = 3$ Gy, at least 800 cc should receive less than 23.4 Gy.

Benchmark case

For the benchmark case, data from a 64-year-old male patient, treated at Erasmus MC, was used. Initially, he had an HCC in segment V-VI, for which he underwent surgery. Three years later, there was a 4 cm lesion visible on CT, compatible with HCC, close to the right branch of the portal vein. There were no enlarged lymph nodes, nor lung metastases. The AFP blood level was 3. MRI showed a 4 cm lesion with wash-out surrounded by a capsule, compatible with HCC. The tumor was located in segment V, transition to VIII and toward segment IV.

In the delineation round, participating centers were asked to define the target (GTV), and delineate the liver for this patient. Arterial and venous phase contrast CT scans (with 2.5 mm slice separation), diagnostic MR images (with 2 mm and 7.7 mm slice separation), and a description of the relevant medical history of the patient were provided. The centers were asked to delineate the liver and GTV according to the trial protocol. The contours were evaluated based on consensus among the medical doctors in the QA team, supported by a radiologist for the GTV. Quantitative comparisons of 3D volumes generated from submitted and reference contours, as defined by the QA team, were based on; the volume difference, the average distance in 3D, the center-of-mass (COM) shift, and the Dice Similarity Coefficient (DSC). The DSC of two volumes is defined as the overlap, divided by the average volume.

In the subsequent planning round, the centers generated for the same patient a treatment plan according to the study protocol, based on contours used for the initial treatment of the patient at Erasmus MC. The results of the treatment planning for the benchmark case were primarily evaluated based on the treatment planning protocol, see the dosimetric parameters listed in Table 1. Also, correlations between the achieved target dose and the NTCP and OAR doses were investigated.

Automated treatment planning

Based on the trial constraints and objectives, fully automated plan generation for HCC was implemented in the in-house developed Erasmus-iCycle software for fully automatic, prioritized
multi-criteria optimization. Erasmus-iCycle was used to generate an individualized planning template for fully automatic generation of a clinically deliverable coplanar VMAT plan with the Monaco TPS (Elekta AB, Stockholm, Sweden) [19,20]. AutoVMAT was primarily used to assess the quality of treatment planning, both for the benchmark case and for clinical trial patients. To this end, fully automatically generated co-planar VMAT plans were compared to 20 treatment plans for 9 patients; 12 for the benchmark case (1 patient, see below), 5 from a clinical pilot study (5 patients), and 3 from the first 3 patients included in the trial so far. Auto-VMAT plans were scaled to the clinical target dose (PTV-D95%, dose to 95% of the PTV). The target dose for the benchmark case and one of the trial patients was 54 Gy in six fractions. For the other 7 patients, it was around 48 Gy, also in six fractions.

### Table 1

<table>
<thead>
<tr>
<th>Item</th>
<th>Protocol</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>III’</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>IX</th>
<th>X</th>
<th>Auto</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality</td>
<td></td>
<td>vmat</td>
<td>vmat</td>
<td>imrt</td>
<td>vmat</td>
<td>ck</td>
<td>vmat</td>
<td>ck</td>
<td>vmat</td>
<td>vmat</td>
<td>vmat</td>
<td>vmat</td>
<td>vmat</td>
</tr>
<tr>
<td>Energy [MV]</td>
<td></td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td>10</td>
<td>6</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Beams/Arcs</td>
<td></td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>MUs</td>
<td></td>
<td>1775</td>
<td>1640</td>
<td>1552</td>
<td>1889</td>
<td>1413</td>
<td>1459</td>
<td>1320</td>
<td>1778</td>
<td>2489</td>
<td>3236</td>
<td>1938</td>
<td></td>
</tr>
<tr>
<td>D_{max} PTV [Gy]</td>
<td></td>
<td>&lt;72</td>
<td>70.5</td>
<td>65.6</td>
<td>67.5</td>
<td>69.5</td>
<td>70.1</td>
<td>65.0</td>
<td>68.5</td>
<td>71.6</td>
<td>68.5</td>
<td>69.6</td>
<td>70.3</td>
</tr>
<tr>
<td>V_{PTV} ≥ 48 Gy [%]</td>
<td></td>
<td>&gt;95</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.3</td>
<td>100</td>
<td>99.6</td>
<td>100</td>
</tr>
<tr>
<td>V_{PTV} ≥ 54 Gy [%]</td>
<td></td>
<td>99.6</td>
<td>99.6</td>
<td>99.6</td>
<td>98.7</td>
<td>99.4</td>
<td>98.4</td>
<td>99.3</td>
<td>100</td>
<td>98.4</td>
<td>100</td>
<td>93.5</td>
<td>100</td>
</tr>
<tr>
<td>V_{Dmax/PTV} ≥ 23.4 Gy [cc]</td>
<td></td>
<td>≤22</td>
<td>925</td>
<td>924</td>
<td>897</td>
<td>965</td>
<td>962</td>
<td>934</td>
<td>926</td>
<td>856</td>
<td>948</td>
<td>904</td>
<td>955</td>
</tr>
<tr>
<td>NTCP_{Low−CTV} [%]</td>
<td></td>
<td>≤5</td>
<td>2.4</td>
<td>3.2</td>
<td>7.0</td>
<td>0.8</td>
<td>2.1</td>
<td>1.2</td>
<td>3.8</td>
<td>22.2</td>
<td>0.2</td>
<td>4.4</td>
<td>0.5</td>
</tr>
<tr>
<td>D_{max} Stomach/ Bowel [Gy]</td>
<td></td>
<td>&lt;39</td>
<td>16.2</td>
<td>16.3</td>
<td>21.4</td>
<td>13.9</td>
<td>18.6</td>
<td>16.4</td>
<td>17.4</td>
<td>18.0</td>
<td>14.3</td>
<td>18.3</td>
<td>15.2</td>
</tr>
<tr>
<td>V_{max} Stomach/ Bowel [cc]</td>
<td></td>
<td>&lt;5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D_{max} Esophagus [Gy]</td>
<td></td>
<td>≤36</td>
<td>1.5</td>
<td>1.9</td>
<td>1.3</td>
<td>1.5</td>
<td>1.7</td>
<td>7.3</td>
<td>4.2</td>
<td>3.3</td>
<td>1.4</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>D_{max} Spinal Cord [Gy]</td>
<td></td>
<td>≤24</td>
<td>6.7</td>
<td>8.7</td>
<td>14.7</td>
<td>11.8</td>
<td>5.1</td>
<td>7.0</td>
<td>6.2</td>
<td>6.0</td>
<td>5.5</td>
<td>8.6</td>
<td>7.4</td>
</tr>
<tr>
<td>D_{max} Right Kidney [Gy]</td>
<td></td>
<td>≤19.2</td>
<td>0.9</td>
<td>0.9</td>
<td>3.8</td>
<td>1</td>
<td>5.4</td>
<td>0.7</td>
<td>3.6</td>
<td>1.4</td>
<td>3.8</td>
<td>2</td>
<td>5.7</td>
</tr>
<tr>
<td>D_{max} Left Kidney [Gy]</td>
<td></td>
<td>≤19.2</td>
<td>2.9</td>
<td>3.7</td>
<td>0.5</td>
<td>2.7</td>
<td>2.9</td>
<td>3.0</td>
<td>1.6</td>
<td>3.2</td>
<td>3.8</td>
<td>2.8</td>
<td>5.7</td>
</tr>
<tr>
<td>Avoidable hotspots</td>
<td></td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
Dose escalation with autoVMAT

Since the trial protocol recommends to cover as much of the PTV as possible with 54 Gy in six fractions, in addition, autoVMAT was also used to assess for how many patients dose escalation to 54 Gy was possible without violating any planning constraint.

Results

Benchmark case – GTV and liver delineation

All ten participating radiotherapy centers completed the contouring. Two institutes primarily contoured on the venous instead of the arterial contrast phase of the CT. All contours were considered acceptable, although variation in both GTV and liver contours was statistically significant (see below). Two axial CT slices with submitted contours are shown in Fig. 1, where also the reference contours are displayed. Several centers delineated parts of the gallbladder, vena cava and diaphragm as liver.

The results of the quantitative comparisons with the reference contours, are shown in Fig. 2. Submitted liver volumes are on average 101.4 cc = 8.0% larger than the reference (range 2.2–12.0%, \( p < 0.001 \)), while submitted GTVs are on average 12.6 cc = 24% smaller than the reference (range 2.8% to 43.2%, \( p < 0.001 \)). The average distance in 3D between submitted and reference volumes ranges

![Graph showing differences between submitted contours and reference contours for all centers, excluding Erasmus MC.](image1)

![Graph showing average 3D distance Liver/GTV to reference for all centers.](image2)

![Graph showing Dice Similarity Coefficients Liver/GTV for all centers.](image3)

![Graph showing Dose escalation with autoVMAT](image4)

Fig. 2. Differences between submitted contours and reference contours for all centers, excluding Erasmus MC. Liver: orange, GTV: purple (COM = center of mass). Distances are based on 3D volumes and the DSC of two volumes is defined as the overlap, divided by the average volume. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
from 1.7 to 3.5 mm for the liver and from 1.2 to 2.4 mm for the GTV. Center-of-mass shifts range from 0.8 to 2.4 mm and from 1.7 to 4.6 mm respectively. DSCs of submitted and reference volumes range from 0.925 to 0.954 for the liver and from 0.721 to 0.876 for the GTV.

**Benchmark case – treatment planning**

Treatment planning results are summarized in Table 1. Two centers (III and VII) did not meet the NTCP constraint initially and re-planned the benchmark case after feedback had been provided. The revised plans are labeled III and VII respectively. For reference, also an autoVMAT plan with a target dose of 54 Gy to 95% of the PTV was generated, see Table 1.

PTV dose homogeneity and conformity vary, with some institutes aiming at a high target dose allowing for large dose gradients in the GTV-PTV margin, and others optimizing for a smoother, more homogeneous, dose distribution. Within the set of 12 plans for the benchmark case, all based on the same CT and structure set, there is little correlation between PTV-D95% and the achieved NTCP value ($R^2 = 0.3, R^2 < 0.1$ for the 10 protocol compliant plans), and hardly any between PTV-D95% and OAR doses ($R^2 < 0.2$ for esophagus, stomach, duodenum and bowel), see Fig. 3a and b. This suggests that the observed differences in dose distribution, especially those within the subset of protocol compliant plans, are largely due to variations in treatment planning and treatment technique.

**Plan QA with autoVMAT**

Within the set of 20 plans, two major and two minor deviations were detected, all with respect to the NTCP. With autoVMAT, one major NTCP deviation of 22.2% was reduced to a minor deviation of 7.5%, while the other was reduced from 28.7% to 12.5%, which is still not protocol compliant. For both these plans, the deviation was caused in part by the dose to the PTV being higher than feasible within the NTCP constraint. The autoVMAT plans could be scaled to fully compliant plans with NTCP values of 0.9% and 3.8%, respectively, by reducing the PTV-D95% from 58.8 Gy to 54.0 Gy and from 50.6 Gy to 48.0 Gy. The major NTCP deviations

![Fig. 3 (continued)](image_url)

**Fig. 3** (continued)

![Fig. 4](image_url)

**Fig. 4.** a: Liver-NTCP for submitted treatment plans versus autoVMAT. Minor protocol deviations are represented in orange, major deviations in red. b: Gastrointestinal OAR doses for submitted treatment plans versus autoVMAT.
were detected in one plan from the benchmark case and one from the clinical pilot. No patients were treated with these plans. In addition, with autoVMAT, two minor NTCP deviations of respectively 8.3% and 7.0%, one in another plan from the benchmark case and one in another plan from the clinical pilot, could be reduced to protocol compliant values of 2.5% and 3.8%, respectively, with the same PTV-D95%.

Although some plans using substantially more modulation/monitor units (MUs) and some non-coplanar (Cyberknife) plans outperform the corresponding co-planar autoVMAT plan, in most cases, autoVMAT resulted in a superior plan, with an average NTCP reduction of 2.2 percent point (range: 16.2 percent point to –1.8 percent point, p = 0.03), a lower mean dose to the healthy liver (p < 0.01) and lower doses to gastrointestinal OARs (p < 0.001), see Fig. 4a and b. For 7 of the 20 plans, including 4 with a protocol deviation, autoVMAT achieved a reduction of the liver NTCP by more than 2 percent point without compromising any other OAR. A reduction of the gastrointestinal (GI) maximum dose(s) of more than 5.0 Gy was achieved for 4 additional plans, without sacrificing the NTCP or increasing the dose delivered to other OARs.

Potential for dose escalation analyzed with autoVMAT

For 7 out of the 9 patients considered here, the clinical target dose was around 48 Gy in six fractions. Using autoVMAT, we investigated if dose escalation to 54 Gy in six fractions was possible within planning constraints for the OARs. For one patient, a target dose larger than 48 Gy was not possible due to an overlap between the PTV and part of the bowel. For another patient, it was not possible to achieve more than 48 Gy without violating the NTCP constraint. For the other 5 patients that were clinically treated to 48 Gy, dose escalation to 54 Gy could be achieved with autoVMAT. In practice, however, clinical considerations other than what can be achieved in treatment planning play a role in deciding on target dose.

Discussion

As part of the TRENDY multi-center randomized trial, an extensive QA program has been implemented, including a benchmark case and prospective (prior to treatment delivery) and retrospective feedback on delineation and treatment planning. Prospective and retrospective feedback during patient accrual promotes the quality and uniformity of SBRT treatment in the trial, but has the possible disadvantage of the trial being less representative for normal clinical practice, especially during the start-up phase when revision of delineation and treatment planning is more likely to be required. Since both the trial protocols, and all QA results are available, other centers should, however, be able to reproduce all (future) trial results.

Submitted benchmark-case delineations were evaluated by comparison with a reference, based on consensus within an expert panel. The submitted delineations of liver and GTV are small in volume compared to the total volume represented by the CT scan, in which case the DSC becomes comparable to kappa values for inter-observer variation in delineation (see supplementary material), as reported for HCC in [21,24,25]. Based on the widely accepted and used interpretation of kappa values in inter-observer studies [26,27], we conclude that there is almost perfect agreement (DSC between 0.8 and 1.0) between all submitted liver contour sets and six of nine GTV contour sets, and that there is substantial agreement (DSC between 0.6 and 0.8) between the other three GTV contour sets and the reference. The DSCs for the liver contours are comparable in magnitude to the kappa values reported in inter-observer studies [21,24,25]. Both individual feedback and general recommendations, based on the observed delineation variations, have been provided to participating centers.

Feedback on treatment planning was based both on the trial protocol and results obtained with fully automated treatment planning. The added value of autoVMAT for plan QA is twofold. It contributes to understanding the nature of protocol deviations in treatment planning, which are often due to suboptimal planning combined with (too) much priority being given to target dose, or to some OAR(s) at the cost of others. In such cases, dosimetric autoVMAT results provide a solution approach, thereby reducing both the number of deviations and patient loss in the trial. In addition, autoVMAT results can be used to identify, quantify and monitor variations in protocol compliant plans, even for an individual patient. Although replanning is not mandatory in such cases, this contributes to the objective and systematic evaluation of submit-
ted treatment plans. AutoVMAT results show what can be achieved in treatment planning, and individual feedback based (in part) on these results, may promote the quality and uniformity of treatment planning for future patients from the same center.

When more patients from multiple centers are included in the trial, automated treatment planning will be applied to identify systematic variations in treatment planning between participating centers and to monitor improvements in treatment planning. Although it is expected that variation in submitted dose distributions is mostly due to treatment planning, at this point it is, strictly speaking, not possible to separately study the impact of treatment planning and treatment technique. In the longer run, dedicated fully automated treatment planning adjusted to treatment delivery techniques in participating centers could be implemented as a part of clinical trials.

As has been demonstrated previously, knowledge-based [3] and automated treatment planning [28] may, in retrospect, reveal suboptimal treatment planning in clinical trials. To the best of our knowledge, the present manuscript constitutes the first report on the use of automated treatment planning for prospective and retrospective plan QA within a clinical trial (while patients are accrued). We have demonstrated that, because of the quality and consistency of automatically generated plans, they can assist in evaluating constraint violations in submitted plans (maybe these can be avoided by changing the plan). Automated planning can also identify suboptimal treatment plans that do fulfill all constraints, but with PTV or OAR dose delivery that could be made better with plan adjustments. Due to the automation, fast feedback on plans submitted by participating centers is feasible.

Conflict of interest

None declared.

Acknowledgements

The TRENDY trial is supported by the Dutch Cancer Society (KWF, project number 2014-6973). We thank Coen Hurkmans, Sebastiaan Breedveld, Chrysì Papalazarou, Wilco Schilleman, Yvette Seppenwoolde, András Zolnay, Maarten Dirx, and Joan Penninkhof for support, feedback and discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.radonc.2017.09.007.

References