Article:
Heijmen, B, Voet, P, Fransen, D et al. (19 more authors) (2018) Fully automated, multi-criterial planning for Volumetric Modulated Arc Therapy - An international multi-center validation for prostate cancer. Radiotherapy and Oncology, 128 (2). pp. 343-348. ISSN 0167-8140

https://doi.org/10.1016/j.radonc.2018.06.023

Fully automated, multi-criterial planning for Volumetric Modulated Arc Therapy – an international multi-center validation for prostate cancer

Ben Heijmen[1*], Peter Voet[2], Dennie Fransen[1], Joan Penninkhof[1], Maaike Milder[1], Hafid Akhiat[2], Pierluigi Bonomo[3], Marta Casati[3], Dietmar Georg[4], Gregor Goldner[4], Ann Henry[5], John Lilley[5], Frank Lohr[6], Livia Marrazzo[3], Stefania Pallotta[3], Roberto Pellegrini[2], Yvette Seppenwoolde[4], Gabriele Simontacchi[3], Volker Steil[6], Florian Stieler[6], Stuart Wilson[5], Sebastiaan Breedveld[1]

[1]Erasmus Medical Center Rotterdam Daniel den Hoed Cancer Center, Radiation Oncology, Rotterdam, The Netherlands.
[2]Elekta AB, Elekta, Stockholm, Sweden.
[3]Azienda Ospedaliero-Universitaria Careggi, Radiation Oncology, Florence, Italy.
[4] Medical University of Vienna, Christian Doppler Laboratory for Medical Radiation Physics for Radiation Oncology / AKH Wien, Vienna, Austria.
[5]St James's Institute of Oncology- St James's Hospital, Radiation Oncology, Leeds, United Kingdom.
[6]University Medical Center Mannheim- Heidelberg University, Radiation Oncology, Mannheim, Germany.

*Corresponding author B.J.M. Heijmen

Phone: +31650001719

Fax: +31107041012

E-mail: b.heijmen@erasmusmc.nl

Number of pages: XXX

Number of tables: 3

Number of figures: 1

Number of figures in Supplementary material: 3

Number of tables in Supplementary material: 5

Running head: Multi-center validation of multi-criterial autoplanning

Keywords: automated multi-criterial treatment planning, knowledge-based planning, VMAT, prostate, multi-center validation

Abstract

Background and purpose: Reported plan quality improvements with autoplanning of radiotherapy of the prostate and seminal vesicles are poor. A system for automated multi-criterial planning has been validated for this treatment in a large international multi-center study. The system is configured with training plans using a mechanism that strives for quality improvements relative to those plans.

Material and Methods: Each of the four participating centers included thirty manually generated clinical Volumetric Modulated Arc Therapy prostate plans (manVMAT). Ten plans were used for autoplanning training. The other twenty were compared with an automatically generated plan (autoVMAT). Plan evaluations considered dosimetric plan parameters and blinded side-by-side plan comparisons by clinicians.

Results: With equivalent Planning Target Volume (PTV) $V_{95\%}$, $D_{2\%}$, $D_{98\%,}$ and dose homogeneity autoVMAT was overall superior for rectum with median differences of 3.4 Gy ($p<0.001$) in D_{mean}, 4.0% ($p<0.001$) in V_{60Gy}, and 1.5% ($p=0.001$) in V_{75Gy}, and for bladder D_{mean} (0.9 Gy, $p<0.001$). Also the clinicians' plan comparisons pointed at an overall preference for autoVMAT. Advantages of autoVMAT were highly treatment center- and patient-specific with overall ranges for differences in rectum D_{mean} and V_{60Gy} of [-4,12] Gy and [-2,15]%, respectively.

Conclusion: Observed advantages of autoplanning were clinically relevant and larger than reported in the literature. The latter is likely related to the multi-criterial nature of the applied autoplanning algorithm, with for each center a dedicated configuration that aims at plan improvements relative to its (clinical) training plans. Large variations among patients in differences between manVMAT and autoVMAT point at inconsistencies in manual planning.

Introduction

Already in 1998 Reinstein et al. [1] investigated automation of inverse planning of radiotherapy. For six prostate cancer patients they could generate high quality Intensity-Modulated Radiation Therapy (IMRT) plans using a fixed planning template. Several papers have now reported on systematic comparisons of fully automatically generated clinically deliverable IMRT or Volumetric Modulated Arc Therapy (VMAT) plans with manually generated plans for treatment of the prostate and seminal vesicles [2-13]. In eight of the twelve studies autoplanning was knowledge-based, i.e. using a model that describes relationships between dose and anatomy in patients treated previously using manual planning. In five of these studies organ-at-risk (OAR) Dose Volume Histograms (DVH) were predicted and used to automatically generate plans for test patients [4,5,8,10,13]. In one study the full dose distribution was predicted [7]. In [2] and [3] automated planning was based on the beam geometry, the fluence map, and the constraints and weights of a reference case in a plan database with similar Beam's Eye View contours as the test patient. Methods of autoplanning not based on a prediction model as derived from a database of previously treated plans included automated a priori multi-criterial plan optimization (aprioriMCO) [6], automated iterative constraint adaptation [9], use of a particle swarm optimizer for automated selection of objective function weights, and automated iterative fine-tuning of cost functions [12]. Only in two studies were plans compared for >50 patients. Reported improvements in plan quality with autoplanning were overall modest. The only report on a multi-center comparison of manual- vs. autoplanning for prostate cancer is by Schubert et al. [10], using a DVH prediction model generated in a single center which was validated in six other centers. Automatically generated plans had reductions in rectum and bladder D_{mean} of 0.6 and 0.8 Gy respectively, while for both OARs high doses were worse for autoVMAT.

In this study we have investigated an aprioriMCO autoplanning system [6, 14, 16, 17]. While in a posteriori MCO a Pareto-frontier of plans is upfront generated for selection of the preferred treatment plan by a planner afterwards, in aprioriMCO a single Pareto-optimal plan is directly and automatically generated for each new patient, featuring clinically favorable trade-offs between all treatment goals. Configuration of the algorithm for a treatment site has an intrinsic mechanism for plan quality improvement relative to training plans [6, 16, 17]. In that sense there is a clear difference with knowledge-based planning that focuses on reproducing the plan quality of previously treated patients.

In [6] aprioriMCO autoplanning was tested for prostate cancer by comparison with manual planning performed by the most competent manual planner in the center whose task was to generate the best possible manual plans without any constraint in planning time. Quality differences between manVMAT and autoVMAT were negligible.

This paper describes a large international multi-center validation of aprioriMCO comparing autoVMAT and manVMAT plans for prostate cancer. In contrast to [6] manVMAT plans were generated with routine clinical planning, so not by the best planner without planning time restrictions. The aprioriMCO algorithm was configured for each center separately, aiming at the best autoVMAT quality for the center's treatment approach with an explicit drive to improve on the quality of the training plans. Plan quality evaluations considered dosimetric plan parameters and side-by-side comparisons of plans by treating clinicians who were blinded for the origin of the plans (autoVMAT or manVMAT). In line with general clinical practice, in these clinicians' comparisons all trade-offs in the plans (doses in PTV and OARs, conformity, etc.) are simultaneously considered for an overall judgement. By including a large number of patients from four centers we could investigate in detail differences between centers and patients in the potential of autoplanning.

Materials and methods

Patients and clinical planning

In each of the participating centers in Mannheim, Florence, Leeds and Vienna (referred to by randomly assigned letters A, B, C, and D) 30 anonymized clinical manVMAT plans recently delivered on an Elekta linac (Elekta AB, Stockholm, Sweden), were included in the study. The prostate and seminal vesicles were irradiated. Plans were generated with manual planning using the Monaco Treatment Planning System (TPS) (Elekta AB, Stockholm, Sweden). A single plan was used in A, B and D, while in center C patients were treated with a sequential boost technique. For the latter center, we investigated plans for delivery of the first 60 Gy. In center A, a simultaneous integrated boost technique was used for delivery of 76 Gy to the prostate and 68.8 Gy to the seminal vesicles in 37 fractions. In centers B and D, the prostate and seminal vesicles were treated up to 78 Gy in 39 fractions, and 80 Gy in 40 fractions respectively. In B patients were treated with an endo-rectal balloon.

Autoplanning in the four centers

The Erasmus-iCycle/Monaco system used in this study for aprioriMCO autoplanning has been extensively described before [6, 14-17]. Input for each plan generation is a tumor site specific wish-list containing hard planning constraints and treatment objectives with assigned priorities used to steer the multi-criterial planning. Automatically generated plans were clinically deliverable. The aprioriMCO system and the procedure for wish-list configuration are summarized in the figures in Figs. E1a and E1b of the Supplementary material.

In each center ten randomly selected manVMAT plans ('training' patients) were used for wish-list tuning (see Tables E2a-E2d of the Supplementary material for final wish-lists). The final wish-list was used for autoVMAT plan generation for the remaining 20 'evaluation' patients (open-loop validation). In each center the Monaco (Elekta AB, Stockholm, Sweden) version used for final automatic plan generation was the same as the version used for manual planning. Also the number of arcs, control points, and the minimum segment size were the same for manual- and autoplanning. Details are provided in Table E3a of the Supplementary material.

In each center autoplanning was completely free from human interference and based on the same manually generated contours as used for manual planning.

Automated vs. manual planning – dosimetric plan quality

AutoVMAT and manVMAT plans were compared by assessing differences in PTV $V_{95\%}$ (coverage), $D_{2\%}$ (near-maximum dose), $D_{98\%}$ (near-minimum dose), and $((D_{2\%}-D_{98\%})/(\text{prescribed dose}))*100$ (homogeneity index, HI), rectum D_{mean}, rectum V_{60Gy}, rectum V_{75Gy}, bladder D_{mean}, and bladder V_{65Gy}. Differences in OAR mean doses represent overall unweighted changes in delivered dose. The use of rectum V_{60Gy}, and V_{75Gy} and bladder V_{65Gy} reflect the concern of the clinicians in this study for high doses in these OARs, with an accent on rectum high dose.

The clinicians first assessed the clinical acceptability of all autoVMAT and manVMAT plans separately (for manVMAT: consistency check as these plans had already been clinically approved for delivery). For this purpose the 20 autoVMAT and 20 manVMAT evaluation plans were individually loaded in the TPS. Pseudo-randomization was used to establish the plan order with corrections to guarantee that in between the autoVMAT and manVMAT plan of a patient there were at least two plans of other patients. For the subsequent blinded side-by-side plan comparisons for the 20 evaluation patients scoring was performed using a visual analogue scale (Fig. E3b of the Supplementary material).

In centers A-C, blinded scoring was performed by one clinician while in center D scoring was performed by two clinicians.

Modulation degree, total MU and estimated treatment time

Auto- and manVMAT plans were compared regarding the degree of modulation, the total number of MU and the estimated treatment time. The degree of modulation, as reported by the TPS, was defined

as the sum of the MU of all segments divided by the sum over all segments of ((segment area · segment MU)/ total beam area).

Deliverability of autoVMAT plans

To verify deliverability of generated autoVMAT prostate plans center A performed QA measurements with the Delta^4 system (Scandidos, Uppsala, Sweden).

Statistics

All differences between autoVMAT and manVMAT were evaluated using paired two-sided Wilcoxon signed-rank tests to assess statistical significance ($p<0.05$). All statistical analyses were performed with SPSS version 24.

Results

For the 80 evaluation patients there was no difference between manVMAT and autoVMAT in median PTV $V_{95\%}$ (98.5 vs. 98.4%, p=0.6), median PTV $D_{2\%}$ (79.5 vs. 80.0 Gy, p=0.9), median PTV $D_{98\%}$ (73.3 vs. 73.5 Gy, p=0.9) and median HI (7.7 vs. 8.1%, p=0.8), while rectum dose was substantially reduced with autoVMAT (median reductions/ranges: 3.4/[-3.7,12.2] Gy ($p<0.001$), 4.0/[-2.3,15.0] % ($p<0.001$), and 1.5/[-3.5,6.7] % ($p<0.001$) for D_{mean}, V_{60Gy}, and V_{75Gy}, respectively). There was also a small advantage for autoVMAT in median bladder D_{mean} (0.9/[-10.1,10.0] Gy, $p<0.001$) while the small difference in bladder V_{65Gy} was not significant (-0.4/[-14.7,15.1], p=0.3). Fig. 1 shows frequency histograms for the observed differences.

While overall PTV coverage was equal, two centers had a small but (borderline) significant advantage for autoVMAT (median differences 0.3% and 1.0%). For the other two coverage was higher for manVMAT (0.4% and 0.8%). In all centers rectum D_{mean} was lower for autoVMAT but for center C the difference was only 0.1 Gy (not significant) while for A and D this went up to 5.6 Gy and 8.0 Gy, respectively. In the latter centers this was accompanied by large reductions in rectum V_{60Gy} (4.4% and 8.7%, respectively). All centers showed a small advantage for autoVMAT in median bladder D_{mean} (0.5 Gy-1.5 Gy) which was (borderline) significant in centers A-C. Although overall there was no statistically significant difference in bladder V_{65Gy}, a small (0.7%) but significant advantage for manVMAT was observed in center A. Table 1 presents an overview of all center-specific differences in plan parameters.

Inter-patient variations in advantages of autoplanning were large (Fig.1 and Table 1). Considering all evaluation patients improvements in rectum D_{mean} and V_{60Gy} ranged from -3.7 to 12.2 Gy and from -2.3% to 15.0%, respectively. While in center D these ranges were as large as [-0.2,12.2] Gy and [-2.3,15.0]%, in center B they were only [-0.8,4.9] Gy and [-0.2,5.4]%, respectively.

100% of manVMAT and 98% of autoVMAT plans were considered clinically acceptable. The clinicians' scores in Table 2 reflect the observed overall superiority of autoVMAT in dosimetric plan parameters. Independent of the clinician considered for D in 29 of 80 comparisons autoVMAT was preferred with a high impact difference with manVMAT (last column Table 2). On the other hand, only in 6-9 comparisons the manVMAT plan was considered superior with high impact. Low impact preferences for autoVMAT and manVMAT were similar. Clinicians' preferences for autoplanning were center specific (compare columns 'A', 'B', 'C' and 'D_1/D_2' in Table 2), in line with the observed differences in dosimetric parameters (Table 1). The treating clinicians in centers A and D preferred more frequently autoVMAT than their colleagues in centers B and C, following the large rectum dose reductions with autoVMAT in the former centers (Table 1).

With autoVMAT the median modulation degree increased from 2.6 to 3.4 with median increases of 13% in MU and 6% in treatment time (Table 3). Larger reductions in rectum D_{mean} and rectum V_{60Gy} were accompanied by larger increases in MU; $p<0.001$ and $R^2=0.3$ for D_{mean}, and $p=0.001$ and $R^2=0.2$ for V_{60Gy}. In center A the modulation degree was considerably higher than in the other centers due to the use of flattening-filter-free treatments.

QA measurements with the Delta^4 system showed a minimum pass rate for autoVMAT plans as high as 98.9%, with clinical acceptance criteria: 90% pass for 3%/3mm.

Discussion

In this large multi-center validation of aprioriMCO for prostate cancer autoVMAT plans were compared to plans that were manually generated in clinical routine in four European centers. There was an overall preference for autoVMAT with important treatment center- and patient-specific variations in the gain with autoVMAT, the latter pointing at inconsistencies in manual planning. Figs. 1b and 1c imply that (large) reductions in rectum D_{mean} with autoVMAT were not at the cost of important increases in rectum V_{60Gy}. On the contrary, reductions in D_{mean} were accompanied with reductions in V_{60Gy} (linear regression analysis; $p<0.001$, $R^2=0.59$).

In centers A, B, and D, large reductions in rectum doses with autoVMAT were observed. In these centers the prescribed tumor dose was ≥ 75 Gy and wish-list configuration was highly focused on reducing (high) rectum dose with autoplanning. Center C did not show significant reductions in

rectum dose, probably related to the relatively low prescribed tumor dose (60 Gy) for the investigated treatment plans.

Although centers A, B and D had similar high PTV prescription doses, the advantage of autoVMAT as expressed both in plan parameters (Table 1) and clinician scoring (Table 2) was clearly highest for A and D. Possibly the planners in center B were often able to manually generate plans close to optimality for the local treatment tradition. Besides center B had very strict PTV coverage restrictions; even a small reduction in coverage to obtain a large gain in rectum dose was often not considered advantageous, reducing the potential gain of autoplanning. Other possible explanations for the lower impact of auto-planning in center B could be the use of a rectal balloon, or suboptimal wish-list tuning in the center.

The two autoVMAT plans that were not clinically acceptable had too high bowel dose, related to absence of bowel delineations other than rectum, prohibiting bowel dose optimization with the inverse planning. A solution could be to delineate for all patients bowel parts potentially at risk for high dose. Alternatively, in the rare cases resulting in unacceptable bowel dose, manual fine-tuning of the autoVMAT plan, based on an added bowel contour, could be used to reduce bowel dose. None of these two approaches were investigated in this study. The two clinically unacceptable autoVMAT plans were in the group of autoVMAT plans that were scored inferior compared to manVMAT, with considered high impact (last column Table 2). In the other cases with high impact advantage for manVMAT, a reduction in OAR dose (generally rectum) with autoVMAT was not considered high enough to justify a (still acceptable) loss in PTV coverage.

The observed increases in modulation degree, MU, and treatment time with autoVMAT in centers A, B, and D would also have been favored in manual planning if there would have been similar quality improvements. Apparently the manual planning processes did not result in this enhanced quality. In center C the considerable increase in MU did not result in increased plan quality. No explanation was found for this observation. It was not considered an important problem in the center as the increase in treatment time (11 sec on 146 sec) was small.

At Erasmus MC Cancer Institute automated plan generation with the investigated aprioriMCO system is in clinical routine since 2013, starting with head and neck cancer (300 patients in 2013). The system is now in routine clinical use for head and neck, prostate, cervix, and advanced lung cancer. QA measurements with the Octavius phantom (PTW, Freiburg, Germany) with γ-evaluation acceptance criteria 90% pass for 3%/3mm do occasionally point at deliverability issues for head and neck cancer patients, not for prostate. In a recent paper on liver SBRT we have reported very high passing rates for plans with high degrees of modulation (4.0 ± 0.9) using the same aprioriMCO system [18]. The dosimetric measurements performed in this study confirmed deliverability of

automatically generated prostate plans. However, certainty on deliverability for future individual patients in a clinical setting requires measurements for each patient independent of the system used for plan generation.

In all published studies on autoplanning for prostate cancer [2,13], differences between automatically and manually generated plans were small, mostly showing advantages of autoplanning in some parameters and disadvantages for others. The overall reported advantage of autoplanning is the large reduction in planning time. In eight of the twelve reported studies autoplanning was knowledge-based [2,3,4,5,7,8,10,13]. In these studies the small differences in plan quality might be related to the rather direct link of automatically generated plans with manually generated training plans in a busy clinical routine. Moreover, with knowledge-based systems there is no mechanism to ensure that final plans are (close to) Pareto-optimal. Also in the [9,11,12], Pareto-optimality was not guaranteed. Detailed comparison of the performed studies is challenging because of differences in set-up: closed-loop vs. open-loop, manual planning in clinical routine or by an expert planner (without time constraints), autoplanning driven by a configuration developed with plans of the same institute or of a different institute, large differences in prescribed tumor dose (50-80 Gy), participating centers may or may not be all academic, large differences in planning goals, constraints and evaluation parameters, and large differences in included patients (only 2 published studies had >50 patients included).

As described in the Introduction section the previously reported multi-center autoplanning validation for prostate cancer by Schubert et al. [10] showed very small differences between auto- and manual planning. The design of that study was very different from our set-up. They used a knowledge-based autoplanning system that has no inherent drive to exceed the quality of the best manVMAT plans used for training, opposed to the aprioriMCO approach used in our study. Their DVH-prediction model for autoplanning was generated in a single center and used for autoplanning in six other centers. In contrast in our study separate autoplanning configurations were performed for each center to maximally exploit the potential of autoplanning for that center. Another difference between [10] and the current study is in the nature of participating centers. While in the former study both academic institutes and departments of regional community hospitals or of private networks of hospitals participated, in our study only academic centers were included. As for all published studies, a limitation of this study was that all manual plans were generated with a single TPS. Possibly plan quality differences could be different for manual plans generated with a different TPS.

This large multi-center study has demonstrated overall superiority of autoVMAT prostate plans compared to manVMAT plans generated in clinical routine. Large variations between patients in the gain of autoVMAT pointed at inconsistencies in manual planning. Superiority of autoVMAT was demonstrated with dosimetric plan parameters and with blinded side-by-side plan comparisons by

clinicians, the latter simultaneously considering all trade-offs in the two plans. The observed gain with autoplanning for prostate cancer was overall larger than reported in the literature. More research is needed to explore the full potential of autoplanning and its optimal clinical application.

295

Acknowledgements

Elekta AB, Stockholm, Sweden is acknowledged for partially funding the investigations. Three co-authors (PV, AA, and RP) are employed by Elekta and had an advisory role in the study design. They were also partly involved in data collection and analysis and had advisory roles in the interpretation of
300 data. PV contributed to the writing of the manuscript. None of the company employees was involved in the decision to submit for publication. We gratefully acknowledge the financial support by the Austrian Federal Ministry of Science, Research and Economy and the Austrian Foundation for Research, Technology, and Development.

305

References

[1] Reinstein LE, Wang XH, Burman CM ,Chen Z, Mohan R, Kutcher G et al. A feasibility study of automated inverse treatment planning for cancer of the prostate. Int J Radiat Oncol Biol Phys. 1998;40:207-14.

[2] Chanyavanich V, Das SK, Lee WR, Lo JY. Knowledge-based IMRT treatment planning for prostate cancer. Med Phys. 2011;38:2515-22.

[3] Good D, Lo J, Lee WR, Wu QJ, Yin FF, Das SK. A knowledge-based approach to improving and homogenizing intensity modulated radiation therapy planning quality among treatment centers: an example application to prostate cancer planning. Int J Radiat Oncol Biol Phys. 2013;87:176-81.

[4] Yang Y, Ford EC, Wu B, Pinkawa M, van Triest B, Campbell P et al. An overlap-volume-histogram based method for rectal dose prediction and automated treatment planning in the external beam prostate radiotherapy following hydrogel injection. Med Phys. 2013;40:011709.

[5] Fogliata A, Belosi F, Clivio A, Navarria P, Nicolini G, Scorsetti M et al. On the pre-clinical validation of a commercial model-based optimisation engine: application to volumetric modulated arc therapy for patients with lung or prostate cancer. Radiother Oncol 2014;113:385-91.

[6] Voet PWJ, Dirkx MLP, Breedveld S, Al-Mamgani A, Incrocci L, Heijmen BJM. Fully automated volumetric modulated arc therapy plan generation for prostate cancer patients. Int J Radiat Oncol Biol Phys 2014;88:1175-79.

[7] Nwankwo O, Mekdash H, Sihono D, Wenz F, Glatting G. Knowledge-based radiation therapy (KBRT) treatment planning versus planning by experts: validation of a KBRT algorithm for prostate cancer treatment planning. Radiat Oncol. 2015;10:111.

[8] Hussein M, South CP, Barry MA, Adams EJ, Jordan TJ, Stewart AJ et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. Radiother Oncol 2016;120:473-9.

[9] Winkel D, Bol GH, van Asselen B, Hes J, Scholten V, Kerkmeijer LG, et al. Development and clinical introduction of automated radiotherapy treatment planning for prostate cancer. Phys Med Biol. 2016;61:8587-95.

[10] Schubert C, Waletzko O, Weiss C, Voelzke D, Toperim S, Roeser A et al. Intercenter validation of a knowledge based model for automated planning of volumetric modulated arc therapy for prostate cancer. The experience of the German RapidPlan Consortium. PLoS One. 2017;12:e0178034.

[11] Yang J, Zhang P, Zhang L, Shu H, Li B, Gui Z. Particle swarm optimizer for weighting factor selection in intensity modulated radiation therapy optimization algorithms. Physica Medica 2017;33:136–45.

[12] Nawa K, Haga A, Nomoto A, Sarmiento RA, Shiraishi K, Yamashita H et al. Evaluation of a commercial automatic treatment planning system for prostate cancers. Med Dosim. 2017;42:203-209.

[13] Kubo K, Monzen H, Ishii K, Tamura M, Kawamorita R, Sumida I et al. Dosimetric comparison of RapidPlan and manually optimized plans in volumetric modulated arc therapy for prostate cancer. Phys. Med. 2017;44:199-204.

[14] Breedveld S, Storchi PRM, Voet PWJ and Heijmen BJM. iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. Med Phys 2012;39:951-63.

[15] Voet PW, Dirkx ML, Breedveld S, Fransen D, Levendag PC, Heijmen BJM. Toward Fully Automated Multicriterial Plan Generation: A Prospective Clinical Study. Int J Radiat Oncol Biol Phys. 2013;85:866-72.

[16] Sharfo AWM, Voet PWJ, Breedveld S, Mens JW, Hoogeman MS, Heijmen BJM. Comparison of VMAT and IMRT strategies for cervical cancer patients using automated planning. Radiother Oncol 2015;114:395-401.

[17] Della Gala G, Dirkx MLP, Hoekstra N, Fransen D, Lanconelli N, van de Pol M et al. Fully automated VMAT treatment planning for advanced stage non-small cell lung cancer patients. Strahlenther Onkol. 2017;193:402-9.

[18] Sharfo AW, Dirkx ML, Breedveld S, Romero AM, Heijmen BJM. VMAT plus a few computer-optimized non-coplanar IMRT beams (VMAT+) tested for liver SBRT. Radiother Oncol. 2017;123:49-56.

Figure legends

Fig. 1. Histograms showing the frequencies of observed differences between manual planning (manVMAT) and autoplanning (autoVMAT) in PTV $V_{95\%}$ (a.), rectum D_{mean} (b.), rectum V_{60Gy} (c.), rectum V_{75Gy} (d.), bladder D_{mean} (e.) and bladder V_{65Gy} (f.). A, B, C, and D represent the participating centers. Not all dosimetric parameters were relevant for all centers because of differences in prescription doses (see text).