Automatic genetic planning for volumetric modulated arc therapy: A large multi-centre validation for prostate cancer

Automatic genetic planning for Volumetric Modulated Arc Therapy: a large multi-centre validation for prostate cancer.

Christian Fiandra[1], Linda Rossi[2], Alessandro Alparone[3], Stefania Zara[3], Anna Sardo[4], Sara Bartoncini[5], Gianfranco Loi[6], Medico Novara[7], Eva Gino[8], Maria Grazia Ruoredda[9], Gian Marco Deotto[10], Medico Siena[11], Stefania Comi[12], Medico IEO[13], Gianluca Ametrano[14], Medico Napoli[15], Lidia Strigari[16], Medico IFO[17], Alessandro Savini[18], Medico Meldola[19], Serena Cargnel[20], Medico Mestre[21], Nunzia Ciscognetti[22], Medico Savona[23], Gabriella Balestra[24], Umberto Ricardi[25] and Ben Heijmen[2].

1. University of Turin, Department of Oncology, Turin, Italy; School of Bioengineering and Medical-Surgical Sciences, Politecnico di Torino, Turin, Italy.
2. Department of Radiation Oncology, Erasmus MC Cancer Institute, Rotterdam, the Netherlands.
3. Tecnologie Avanzate, Torino, Italy
4. Medical Physics Unit, A.O.U. Città della Salute e della Scienza, Corso Bramante 88/90, 10126 Turin, Italy.
5. Radiation Oncology, Department of Oncology, AOU Città della Salute e della Scienza, Turin, Italy.
6. …
7. …
8. …
9. …
10. …
11. …
12. …
13. …
14. …
15. …
16. …
17. …
18. …
19. …
20. …
21. ..
22. ..
23. ..
24. ..
25. Department of Oncology, University of Turin, Turin, Italy

Purpose

The first clinically applied genetic autoplanning algorithm (Genetic Planning Solution, GPS) was validated in ten radiotherapy centres for prostate cancer VMAT by comparison with manual planning (Manual).

Methods

Although there were large differences among centres in planning protocol, GPS was tuned for a single centre and then applied everywhere without any centre-specific fine-tuning. For each centre, ten Manual plans were compared with autoGPS plans, considering dosimetric plan parameters and the Clinical Blind Score (CBS)

resulting from blind clinician plan comparisons. AutoGPS plans were used as is, i.e. there was no patient-specific fine-tuning.

Results

For nine centres, all ten plans were clinically acceptable. In the remaining centre, only one plan was acceptable. For the 91% acceptable plans, differences between Manual and AutoGPS in target coverage were negligible. OAR doses were significantly lower in AutoGPS plans ($p<0.05$); rectum $D_{15\%}$ and D_{mean} were reduced by 8.1% and 17.9%, bladder $D_{25\%}$ and D_{mean} by 5.9% and 10.3%. According to clinicians, 69% of the acceptable AutoGPS plans were superior to the corresponding Manual plan. In case of preferred Manual plans (31%), perceived advantages compared to autoGPS were minor. QA measurements demonstrated that autoGPS plans were deliverable. A quick configuration adjustment in the centre with unacceptable plans rendered 100% of plans acceptable.

Conclusion

A novel, clinically applied genetic autoplanning algorithm was validated in ten centres for in total 100 prostate cancer patients. Using a single algorithm configuration for all centres, autoplans were overall superior to corresponding manually generated plans.

1 INTRODUCTION

Creation of treatment plans is labor intensive, while the quality may depend on planner skills and ambitions, and on allotted planning time [1,2]. Planning automation has been suggested for overcoming these issues. Hussein et al. have recently reviewed the current literature for IMRT/VMAT auto-planning [3]. The autoplanning systems were divided in three groups:

i) Knowledge-based planning (KBP) to predict parameters (e.g. DVHs) for plan generation of a new patient, based on a database of plans of previously treated patients. The basic assumption is that patients with similar anatomies should have similar optimal dose distributions [4-6],
ii) Protocol-based automatic iterative optimization, based on automatic, patient-specific adjustments of cost functions and constraints [7,8],
iii) a posteriori or a priori multicriteria optimization (MCO). MCO always results in a Pareto-optimal plan. In a posteriori MCO [9, 10], a set of Pareto-optimal plans is automatically generated, while manual Pareto navigation is used to select the final, clinically most attractive plan. In a priori MCO, only one Pareto-optimal plan is generated. By using a priori defined criteria in the plan generation, this plan is also clinically favorable [11-14].

For all three groups, commercial implementations are currently available [3].

Common outcome of autoplanning validation studies is the large reduction in planning workload [3]. This has also been observed in studies with manual fine-tuning of autoplans to arrive at the final plan [15]. Depending

on the system and the study, also gain in plan quality is reported. All autoplanning systems need to be configured for generation of desired plans. This labor-intensive task has to be done for each tumor site, and is generally performed in the treating centre. Auto-planning performance has been mostly investigated in single centre studies, but a few multi-centre validation studies have been performed as well [16-23]. The main advantage of multi-centre studies, compared to single centre, is the implicit testing of generalizability of a methodology/configuration that was developed in a single centre.

Recently, we reported on a novel auto-planning algorithm for VMAT/IMRT, based on genetic optimization ('Genetic Planning Solution' (GPS)) [24]. To the best of our knowledge, this was the first paper investigating genetic optimization for VMAT auto-planning. The system is used in clinical routine since the beginning of 2019 at the University hospital Città della Salute e della Scienza in Turin, Italy. Currently it is applied for prostate, rectum, head and neck, and liver and lung SBRT. In Hussein's classification, this is a type ii algorithm. Genetic algorithms are inspired by the principle of natural selection and biological evolution working on a population of potential solutions by applying the principle of survival of the fittest [25]. Genetic optimization is flexible and can avoid getting trapped in local minima, both for convex and non-convex cost functions.

In this paper, we report on a large multi-centre validation of the genetic GPS autoplanning system with ten participating centres, including in total 100 patients. While there was a large variation in treatment protocols in the centres, the algorithm was configured in one centre, and then used for autoplanning in the others, without any centre-specific fine tuning. The hypothesis was that high-quality autoplanning with GPS was feasible without workload intensive, centre-specific configuration. For validation, the autoGPS VMAT plan of each study patient was compared to the corresponding VMAT plan that was manually generated and clinically applied (Manual). Apart from commonly used dosimetric plan comparisons, an important component of the validation were blind clinician plan comparisons, involving ten clinicians, one for each centre. Deliverability of the autoGPS plans was assessed by patient-specific dosimetric QA measurements.

2 Materials and Methods

2.1 Patients and Manual plans

Table E1 summarizes for each of the ten participating centres the main planning requirements for PTV and OARs. In 9/10 centres, patients were treated on regular linacs. In centre D, tomotherapy was used for treatment. In each centre, the Manual plans were generated in the normal clinical routine, i.e. each centre was asked to retrieve from its database ten recent, randomly selected VMAT/tomotherapy plans of patients already treated. Generally, the plans submitted by a centre were generated by more than one planner.

2.2 GPS automated planning

The GPS autoplanning system was described in detail in [22]. Python scripts were used for integrating the algorithm in the RayStation TPS, version 8A (RaySearch, Stockholm, Sweden). For genetic planning, a fitness function (FF) needs to be defined to evaluate solutions and steer the overall process toward generation of a favorable plan. In this study, the FF was tuned for generation of plans in centre A (where the system was developed). The configuration of the GPS for this study is summarized in Paragraph E1 of the electronic appendix.

For plan generation, CT images and contours were first loaded into the TPS. Then, the dose prescription was defined for each Planning Target Volume (PTV) as reported in Table E1 (third column). After this, the final plan was generated without any human interaction, i.e. there was no patient-specific fine-tuning of plans apart from a rescaling if desired.

2.3 Plan evaluations and comparisons

2.3.1 Analyses of treatment plan parameters

As visible in Table E1, there was a large variation among centres in parameters used for plan evaluation. In this study, this inter-centre variability was (implicitly) taking into account in assessments of plan acceptability and blind plan comparisons by clinicians (see below). However, for population-based dosimetric analyses, common parameters were used for all centres, summarizing the many different requirements shown in Table E1. Collected parameters were: PTVs $V_{95\%}$ (coverage), $V_{107\%}$ (overdosage) for the high-dose PTV (PTV_HD) and the low-dose PTV (PTV_LD), conformity index ($CI_{50\%}$), rectum D_{mean}, rectum $D_{15\%}$, bladder D_{mean}, bladder $D_{25\%}$, femoral heads D_{mean}, and penile bulb D_{mean}. For rectum and bladder, percentage volume levels (X, in $D_{X\%}$) were chosen since they make possible to combine data with different dose levels. X=15% and X=25% for rectum and bladder, respectively, were taken from the literature [27-28].

2.3.2 Clinicians' evaluations

Prior to the blind comparisons of autoGPS plans with corresponding Manual plans, the clinician first assessed clinical acceptability of all plans both manual and auto; if the centre-specific dosimetric goals for the PTVs and OARs, as defined in Table E1, were met (condition usually guaranteed by planning physicist), the acceptability was evaluated by visual inspection of the dose distribution. In another session, the clinician performed for all 10 patients a blinded side-by-side comparison of the Manual and AutoGPS plans, loading the double view for isodoses and DVHs. If requested by the clinician, a template of centre-specific clinical goals (Table E1, requirement column) was displayed to support the clinician in the comparison of the plans. For each comparison, the clinician used a visual analog scale (VAS) to score the overall plan quality difference, following an idea by Heijmen et al. [19]. The clinician had to first select which plan was the better-one, and then indicate the importance of the plan quality difference on a scale from 0 to 100, with 3 global importance levels (Low, Medium, High, see Figure 1).

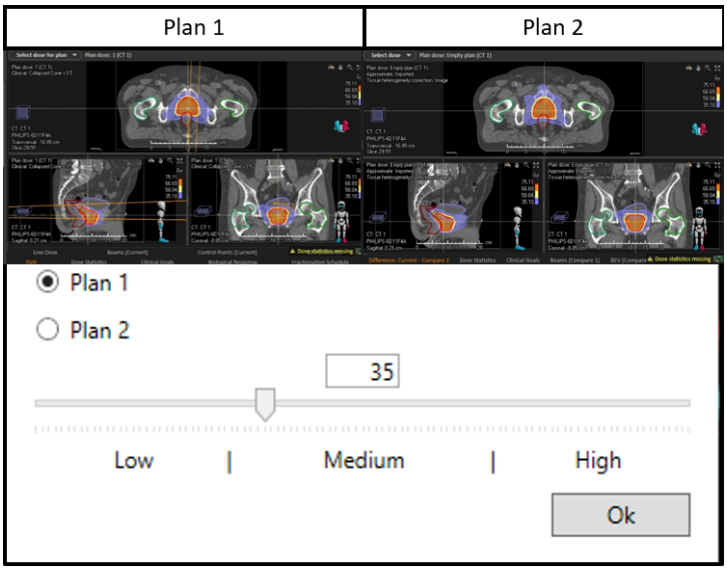


Figure 1.Visual tool for blind plan comparisons by clinicians. The depicted visual analogue scale starts at score 0 on the left and finishes at 100 on the right. For guidance, the scale is divided in three regions expressing perceived small plan quality differences (Low, < 33), intermediate differences (Medium, > 33 and < 66), and large differences (High, > 66). In the example, the clinician selected Plan 1 as the better-one, with an intermediate quality difference, expressed by a Clinical Blind Score (CBS) of 35.

2.4 Deliverability

Deliverability of the AutoGPS plans was verified at Institution A by performing QA measurements with a Delta4 system (Scandidos, Uppsala, Sweden) for all ten study patients, and comparisons with the QA results for the Manual plans.

2.5 Optimization times

The plan optimizations were performed on different hardware configurations: centre A was equipped with double intel Xeon SP 4116, 16.5M Cache, 2.10 GHz with 128 GB ram mounted on a remote server. Optimization times were automatically measured and reported.

2.6 Statistical analyses

Paired two-sided Wilcoxon signed-rank tests were carried out for all dosimetric parameters with 5% significance level ($p<0.05$). MedCalc® version 15.8 was used for all statistical analyses.

3 Results

3.1 AutoGPS plan acceptability

For nine of the ten centres, all ten autoGPS plans were acceptable to the treating clinician, even without re-scaling. In centre H, none of the plans obtained sufficient PTV coverage. With re-scaling, sufficient coverage was obtained for all ten, but this resulted in a too high PTV $V_{107\%}$. For one of the patients, the clinician found the autoplan still clinically acceptable because of low OAR doses. Therefore 91% of autoGPS plans were acceptable. In the Discussion section, an explanation for the 9% not acceptable plans in centre H is discussed, together with a quick solution which is already in place.

3.2 AutoGPS vs Manual – dosimetric plan quality

3.2.1 Combined analyses for all centres and all patients for acceptable plans

The combined analyses are graphically presented in Fig. 2. In the 91 comparisons including acceptable plans, PTV coverage differences between Manual and AutoGPS were negligible (PTV_HD Manual $V_{95\%}$ = 97.9 vs. PTV_HD AutoGPS $V_{95\%}$ = 98.0, p=0.81; PTV_LD Manual $V_{95\%}$ = 98.5 vs. PTV_LD AutoGPS $V_{95\%}$ = 97.8, p=0.13). PTV overdosage was also comparable (PTV_HD Manual $V_{107\%}$ = 0.38 vs. PTV_HD AutoGPS $V_{107\%}$ = 0.36, p=0.65). PTV_LD $CI_{50\%}$ was reduced with AutoGPS by 9.6%, but this was not significant (p=0.52). Apart from penile bulb D_{mean}, all other OAR dose parameters were significantly lower for AutoGPS ($p<0.05$). Rectum $D_{15\%}$ and D_{mean} were reduced by 8.1% and 17.9%, bladder $D_{25\%}$ and D_{mean} by 5.9% and 10.3%, and left and right femoral heads D_{mean} by 6.2% and 6.1% with AutoGPS.

In 63 of 91 (69%) blind comparisons including an acceptable autoGPS plan, the clinician preferred this plan. In case autoGPS was best, the median CBS was 15 (95% CI for the median: 10 to 20). In the 28/91 (31%) comparisons that resulted in a preference for the Manual plan, the median CBS was 10 (95% CI for the median 10 to 20). Monitor Units (MU) were significantly lower for Manual (10.6%; $p<0.001$).

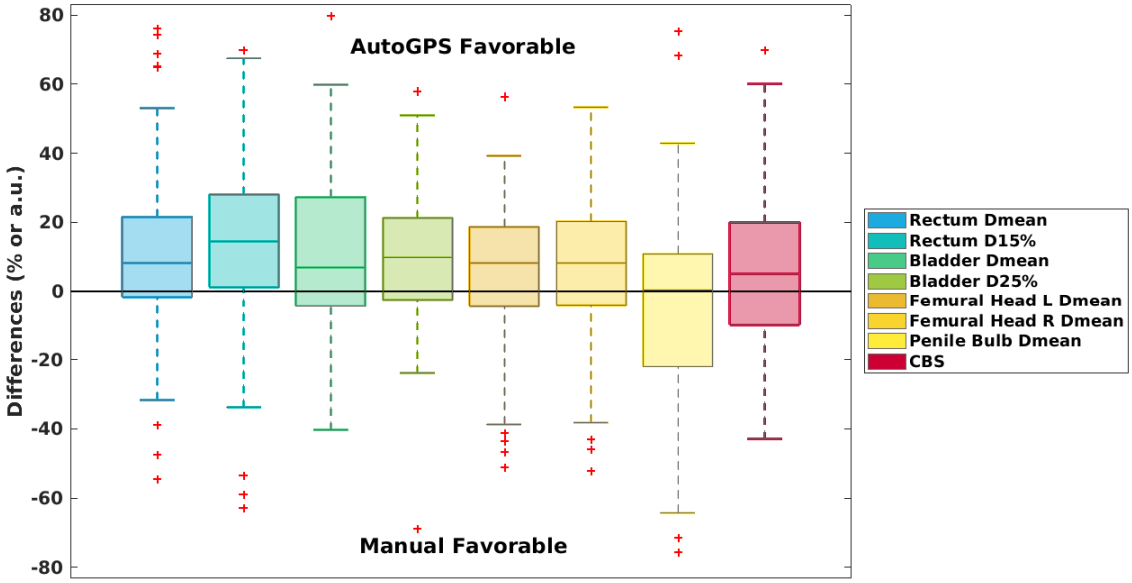


Figure 2. Boxplots showing percentage differences between Manual and AutoGPS for dosimetric OAR parameters. For CBS, the VAS scores obtained with the scoring tool shown in Fig. 1 are presented; scores for blind comparisons resulting in a favorable Manual plan are presented as negative values. All differences were statistically significant in favor of AutoGPS ($p<0.05$), except for D_{mean} of the penile bulb. boxes: 25-75 percentiles, whiskers: 10 to 90 percentiles, crosses: outliers.

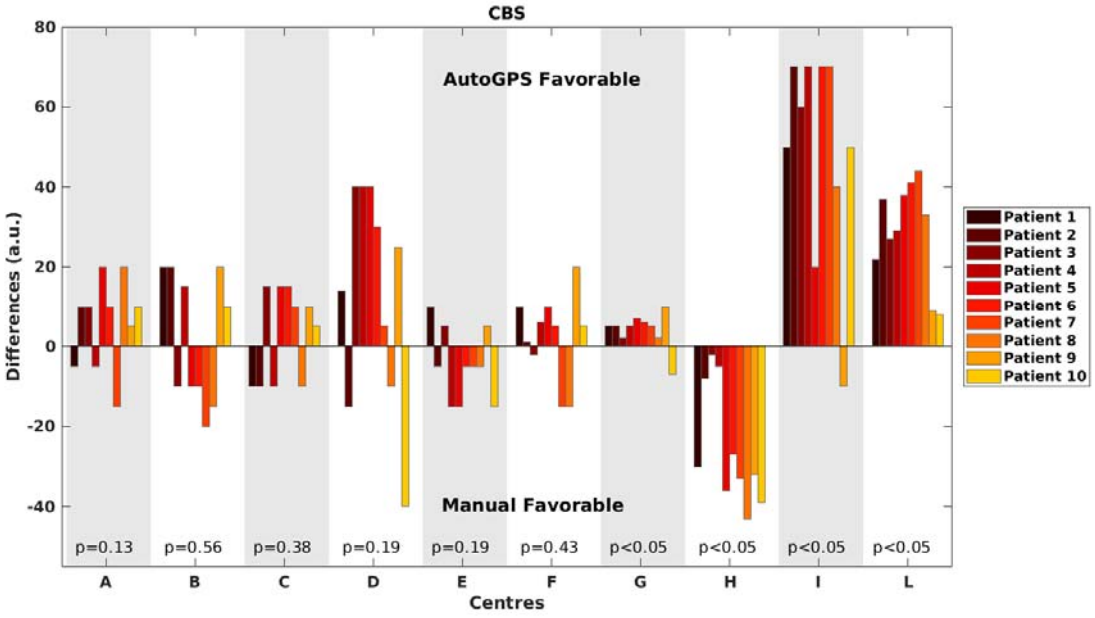


Figure 3. CBS for the ten patients in each of the ten participating centres. For each centre, the ten bars represent a patient. For blind comparisons resulting in a favorable Manual plan, the CBS derived from the VAS (Figure 1) is presented as a negative value. In centre H, only the third patient had a clinically acceptable autoGPS plan. For all other centres, all autoGPS and Manual plans were acceptable.

3.2.2 Per centre and per patient analyses

In Fig. 3 CBS values are presented for each patient, and they are grouped per centre. Apart from centre H with nine unacceptable autoGPS plans, clinicians overall clearly preferred autoGPS. On the other hand, inter-centre variations were clearly visible (compare centres I and L with centres E and F). In three of the nine centres with all autoGPS plans acceptable, autoGPS was statistically significant better, and in the other six centres there was no significant difference. Also, within centres, there were large inter-patient differences in CBS.

Fig.4 presents for the ten participating centres mean differences in autoGPS and Manual dosimetric plan parameters for the ten study patients. Overall, autoGPS was clearly superior, but again there were also large inter-centre variations. As described also above, the patients in centre H had a large overshoot in PTV HD V107%, rendering nine/ten plans unacceptable.

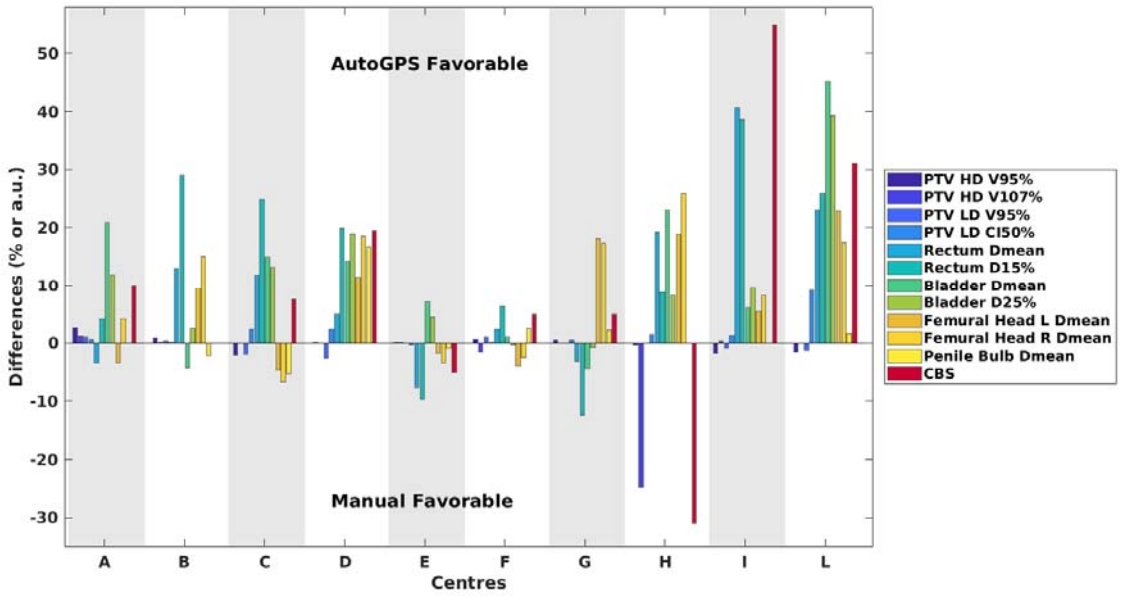


Figure 4. For each centre, mean percentage differences between autoGPS and Manual plan parameters and the mean CBS. Each bar represents a plan parameter/CBS.

3.3 Plan deliverability

QA measurements with the Delta4 system showed for AutoGPS plans a gamma pass rate of 95.2% vs. 97.3% for Manual plans (p=0.52) for clinical acceptance criteria, i.e. 90-95% pass for the 3%/2 mm global gamma criterion [29].

3.4 Optimization times

Mean time for autoGPS plan generations was 20 min, ranging from 15 to 30 min, mainly depending on the available hardware resources.

4 Discussion

The first clinically applied genetic autoplanning algorithm (GPS) was validated for prostate VMAT in a multi-centre setting. For ten patients in each of ten different Italian radiotherapy centres, autoGPS plans were compared to Manual plans based on i) PTV and OAR plan parameters, ii) clinicians' blind plan comparisons (clinician blind score, CBS), iii) delivered MU, and iv) deliverability. Although there was a large variation in applied planning protocols in the participating centres, the GPS was configured in one of the centres and then applied as is for autoplanning in all centres, without further tuning of the configuration. We hypothesized that the applied genetic planning would be flexible enough for high-quality plan generation in all centres, without workload intensive centre-specific configurations. Of the 100 autoGPS plans, 91% was clinically acceptable. For 69% of the patients with an acceptable autoGPS plan, the clinician considered this plan superior to the Manual plan, as expressed by the CBS. The single OAR dose parameters also pointed at an overall advantage for the autoGPS plans. More specifically, rectum and bladder dose parameters were lowest in the autoGPS plans, for equal PTV coverage.

On the other hand, clear inter-centre variations were observed in plan quality gain with autoplanning. This could be related to inter-centre variations in the quality of clinical plans, but it could also be that the single GPS configuration, used in all centres, did better fit with the planning protocol in some centres than with the protocol in others (Table E1 shows large protocol variations). This is a topic for further study. Also, for the observed large inter-patient variations in CBS within centres several explanations are possible, e.g. variations in case complexity, different planners, variations in time spent on planning.

The problems with autoGPS plans observed in centre H (9/10 unacceptable) were due to a minor configuration issue, related to the local PTV coverage criterion that differed substantially from the criterion applied in centre A were the configuration was built. The problem was resolved in a later version of the software, allowing selection of desired coverage (V_{95}>95%, V_{95}>98% or V_{100}>95%). When running automated planning again for centre H with this new option, the PTV coverage differences between Manual and AutoGPS were negligible while overdosages were within prescribed limits.

With excellent autoGPS results for an autoplanning configuration built in centre A in eight other centres, and with a minor configuration adjustment, also in the centre H, there is an indication for good generalizability of configurations of the investigated genetic algorithm. So far it is not clear whether this is related to the applied genetic optimization. Further work is also needed for other tumor sites to confirm the observations.

In a genetic planning algorithm, multiple plan optimizations are sequentially performed to find an optimal solution. The process is controlled by genetic steering. Due to the multiple optimizations, genetic optimization can result in long computation times. However, in this study we have seen that in a regular clinical setting of server-based architecture with high computation power, median optimization times were only 20 min, which we consider clinically acceptable. Performance for other tumor sites is a topic for further study.

With ten participating centres and in total 100 included patients, this is the largest multi-centre validation of autoplanning so far. However, with only ten patients per centre to keep the workload feasible, some caution is needed in the interpretation of presented inter-centre comparisons.

In the literature, (multi-institutional) validation of plan quality in autoplanning is mostly based on dosimetry [3]. Schubert et al. [16] validated a DVH prediction model, generated in a single centre, in seven other centres for in total 60 prostate patients. Heijmen et al. [19] tested a multi-criterial autoplanning algorithm for 4 centres with 20 prostate cancer patients each, using a dedicated configuration for each centre that aimed at plan improvements relative to its (clinical) training plans (10 per centre); clinician scoring was performed. Roach et al [20] investigated whether pre-existing locally developed prostate configurations for a commercial autoplanning TPS were flexible across clinics with different treatment planning protocols, involving three

participating centres using a three-patient training dataset circulated from each centre (10 patients from 3 centres): plan quality was only assessed through DVH analysis and protocol compliance.
Li et al. [21], Kavanaugh et al. [22] and Younge et al. [23] worked on an efficient method for training and validation of a knowledge-based planning (KBP) system as a clinical trial plan quality-control system: DVH metrics and normal tissue complication probabilities were compared.
Fogliata et al. [18] used a set of 70 previously treated esophageal patients in two different institutions to train a model for the prediction of dose-volume constraints, which was then applied in a third centre, not participating in the training. Plan comparison was performed with DVH analysis, complemented with blind clinician comparisons. Ten clinicians were involved in these investigations. Some parameters were better for one technique (e.g. autoplanning) and some other parameters for the other technique (e.g. manual planning). Therefore, an overall score was generally needed for a final verdict. Importantly, this score should also take into account the total dose distribution (conformality etc), which cannot be (easily) summarized in parameters. This overall plan quality scoring is typically a task for treating clinicians, as creating an overall opinion on treatment plans is part of their daily work. Scoring by clinicians is also the only way to answer the question whether autoplans can (and will) be used in clinical routine.

5 Conclusion

A novel, clinically applied genetic autoplanning algorithm was validated in ten centres for in total 100 prostate cancer patients by comparison with manually generated plans. While clinical planning protocols in the centres showed large differences, and only one centre was involved in algorithm configuration, autoplans were overall superior to corresponding manually generated plans, both in terms of dosimetric plan parameters, and scores obtained from blind clinician plan comparisons. Autoplanning with genetic optimization is promising for reducing planning workload, obtaining high plan quality, and avoiding time-consuming centre-specific configuration. Future research for more tumor sites is needed to confirm results.

References

[1] Marino C, Villaggi E, Maggi G, et al. A feasibility dosimetric study on prostate cancer : are we ready for a multicentre clinical trial on SBRT? Strahlenther Onkol. 2015 Jul;191(7):573-81
[2] Nelms BE, Robinson G, Markham J, et al. Variation in external beam treatment plan quality: an inter-institutional study of planners and planning systems. Pract Radiat Oncol. 2012;2:296–305.
[3] Hussein M, Heijmen BJM, Verellen D, Nisbet A Automation in intensity modulated radiotherapy treatment planning-a review of recent innovations. Br J Radiol. 2018 Dec;91(1092):20180270. doi: 10.1259/bjr.20180270. Epub 2018 Sep 4.
[4] Hussein M, South CP, Barry MA, et al. Clinical validation and benchmarking of knowledge-based IMRT and VMAT treatment planning in pelvic anatomy. Radiother Oncol. 2016; 120:473–479.
[5] Kubo K, Monzen H, Ishii K, et al. Dosimetric comparison of Rapid-Plan and manually optimized plans in volumetric modulated arc therapy for prostate cancer. Phys Med. 2017;44:199–204
[6] Prichard, H.; Wo, J.; Wang, Y. Development and Evaluation of a Machine Learning Optimization Model for Automatic Planning in Liver SBRT Using a Multicriteria Optimization Knowledge Base in RayStation. MEDICAL PHYSICS Volume: 46
[7] 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.
Issue: 6 Pages: E417-E418 Meeting Abstract: WE-C930-Ge Published: JUN 2019
[8] 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;42:203–9.
[9] Teichert K, Suss P, Serna JI, Monz M, Kufer KH, Thieke C. Comparative analysis of Pareto surfaces in multi-criteria IMRT planning. *Phys Med Biol* 2011; **56**: 3669–84.
[10]. Monz M, Kufer KH, Bortfeld TR, Thieke C. Pareto navigation—algorithmic foundation of interactive multi-criteria IMRT planning. *Phys Med Biol* 2008; **53**: 985–98.
[11] 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;88:1175–9.
[12]. Breedveld S, Storchi PRM, Keijzer M, Heemink AW, Heijmen BJM. A novelapproach to multi-criteria inverse planning for IMRT. *Phys Med Biol* 2007; **52**: 6339–53.

[13]. Breedveld S, Storchi PRM, Voet PWJ,Heijmen BJM. iCycle: integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. *Med Phys* 2012;**39**: 951–63.
[14]. 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**: 1175–9.
[15]. Vanderstraeten B, Goddeeris B, Vandecasteele K, van Eijkeren M, De Wagter C, Lievens Y Automated Instead of Manual Treatment Planning? A Plan Comparison Based on Dose-Volume Statistics and Clinical Preference. Int J Radiat Oncol Biol Phys. 2018 Oct 1;102(2):443-450. doi: 10.1016/j.ijrobp.2018.05.063. Epub 2018 Jun 2
[16] Schubert C, Waletzko O, Weiss C et al. Intercentre 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.
[17]. Wu B, Kusters M, Kunze-busch M, Dijkema T, McNutt T, Sanguineti G, et al. Crossinstitutional knowledge-based planning (KBP) implementation and its performance comparison to Auto-Planning Engine
Radiother Oncol 2017; 123: 57–62.
[18]. Fogliata A, Nicolini G, Clivio A, Vanetti E, Laksar S, Tozzi A, et al. A broad scope knowledge based model for optimization of VMAT in esophageal cancer: validation and assessment of plan quality among different treatment centres. Radiat Oncol 2015;
[19] Heijmen B, Voet P, Fransen D. Fully automated, multi-criterial planning for Volumetric Modulated Arc Therapy - An international multi-centre validation for prostate cancer. Radiother Oncol. 2018 Aug;128(2):343-348.
[20] Dale Roach, Geert Wortel, Cesar Ochoa. Adapting automated treatment planning configurations across international centres for prostate radiotherapy. Physics and Imaging in Radiation Oncology 10 (2019) 7–13
[21]. Li N, Carmona R, Sirak 2, Kasaova L, Followill D, Michalski J, Bosch W, Straube W, Mell LK, Moore KL. Highly Efficient Training, Refinement, and Validation of a Knowledge-based Planning Quality-Control System for Radiation Therapy Clinical Trials.Int J Radiat Oncol Biol Phys. 2017 Jan 1;97(1):164-172.
[22] Kavanaugh JA, Holler S, DeWees TA, Robinson CG, Bradley JD, Iyengar P, Higgins KA, Mutic S, Olsen LA.
Multi-Institutional Validation of a Knowledge-Based Planning Model for Patients Enrolled in RTOG 0617: Implications for Plan Quality Controls in Cooperative Group Trials. Pract Radiat Oncol. 2019 Mar;9(2):e218-e227. doi: 10.1016/j.prro.2018.11.007. Epub 2018 Dec 15
[23] Younge KC, Marsh RB, Owen D, Geng H, Xiao Y, Spratt DE, Foy J, Suresh K, Wu QJ, Yin FF, Ryu S, Matuszak MM Improving Quality and Consistency in NRG Oncology Radiation Therapy Oncology Group 0631 for Spine Radiosurgery via Knowledge-Based Planning. Int J Radiat Oncol Biol Phys. 2018 Mar 15;100(4):1067-1074. doi: 10.1016/j.ijrobp.2017.12.276. Epub 2018 Jan 4
[24] Fiandra C, AlparoneA. Gallio E et al. Automated Heuristic Optimization of Prostate VMAT Treatment Planning. January 2018 International Journal of Medical Physics, Clinical Engineering and Radiation Oncology 07(03):414-425
[25] Engelbrecht, Andries P. Computational Intelligence: An Introduction. 2007.
[26] Fiorino C, Fellin G, Rancati T, et al. Clinical and dosimetric predictors of late rectal syndrome after 3D-CRT for localized prostate cancer: Preliminary results of a multicentre prospective study. Int J Radiat Oncol Biol Phys 2008;70: 1130–1137
[27] Kuang Y, Wu L, Hirata E et al. Volumetric modulated arc therapy planning for primary prostate cancer with selective intraprostatic boost determined by 18F-choline PET/CT. Int J Radiat Oncol Biol Phys. 2015 Apr 1;91(5):1017-25.
[28] Marks LB, Yorke ED, Jackson A et al. Use of normal tissue complication probability models in the clinic. Int J Radiat Oncol Biol Phys. 2010 Mar 1;76(3 Suppl): S10-9
[29] Miften M., Olch A., Mihailidis D. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys. 2018 Apr;45(4):e53-e83.

.
.

.

Centre	Target	Dose Prescription (Gy)	Fractions	Dose/ fraction (Gy)	Structures	Requirements	TPS for Manual
A	Prostate gland seminal vescicles	61,1	26	2,35	PTV-HD PTV-LD Rectum Bladder	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% $V_{95\%}$≥95-98% V_{67Gy}<15%, V_{58Gy}<25% V_{64Gy}<35%, V_{73Gy}< 15%, V_{70Gy}<25%	Monaco
	Prostate gland	70,2		2,7	Femural head Penis bulb	D_{mean} <44Gy D_{mean} <50Gy	
B	Prostate gland seminal vescicles	56,25	25	2,25	PTV-HD PTV-LD Rectum Bladder	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% $V_{95\%}$≥95-98% V_{35Gy}<45%, V_{53Gy}<35%, V_{61Gy}<20%, V_{65Gy}<10%, $D_{3\%}$<67.5Gy $D_{1\%}$<70Gy V_{45Gy}<80%, V_{57Gy}<50%, V_{61Gy}<35%, $V_{67.5Gy}$<15%	RayStation
	Prostate gland	67,5		2,7	Femural head Penis bulb	V_{40Gy}<2%,D_{max}<45Gy D_{mean} < 50Gy	
C	Prostate gland seminal vescicles	63,6	30	2,12	PTV-HD PTV-LD Rectum Bladder	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% $V_{95\%}$≥95-98% V_{46Gy}<50%, V_{65Gy}<25%, V_{69Gy}<15%, V_{70Gy}<10cc V_{60Gy}<50%,V_{70Gy}<25%, V_{74Gy}<15%	RayStation
	Prostate gland	72		2,4	Femural head Penis bulb	V_{46Gy}<50%, D_{mean}<48Gy V_{42Gy}<60%, D_{mean}<49Gy	
D	Prostate gland seminal vescicles	76,5-61,2	34	2,25-1,8	PTV-HD PTV-LD Rectum Bladder	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% $V_{95\%}$≥95-98% V_{56Gy}<35%, V_{60Gy}<25%, V_{65Gy}<15% V_{55Gy}<50%,V_{60Gy}<30%, V_{70Gy}<25%	Volo
	Prostate gland	70	35	2	Femural head Penis bulb	$V_{46.2Gy}$<5%,D_{max}< 50Gy D_{mean} <50Gy	
E	Prostate gland seminal vescicles	35	5	7	PTV-HD Rectum Bladder Femural head Penis bulb small bowel Penis Testis Anal Canal	$D_{95\%}$≥95%; D_{max}(%)≤111%; $V_{110\%}$<20% $V_{50\%}$ <50% $V_{80\%}$< 20% $V_{90\%}$< 10% $V_{100\%}$< 5% $V_{100\%}$ <10% $V_{50\%}$< 40% $V_{100\%}$< 5cm^3 V_{15Gy}< 5% V_{29Gy}< 50% V_{30Gy}<1cm^3 D_{mean}<15% V_{17Gy}< 195cc V_{13Gy}<1cm^3 $D_{20\%}$<2Gy D_{mean}<15Gy	RayStation
F	Prostate gland seminal vescicles	72-64,5	30	2,4-2,15	PTV-HD PTV-LD Rectum Bladder	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% $V_{95\%}$≥95-98% V_{50Gy} <35%, V_{60Gy}< 25%, V_{70Gy}< 5% V_{65Gy}<35%, V_{70Gy}< 15%	RayStation
	Prostate gland	72		2,4	Femural head Penis bulb	D_{mean}< 20Gy, D_{max}< 50Gy D_{mean}< 50Gy	
G	Prostate gland	62	20	3,1	PTV-HD Rectum Bladder Penis Bulb Femural head	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% V_{32Gy}<50% V_{60Gy}<15%, $D_{0.1cc}$<63.5Gy V_{32Gy}<50% V_{60Gy}<20%, $D_{0.1cc}$<63.5Gy D_{max}<100% $D_{0.1cc}$ <34Gy	Pinnacle
H	Prostate gland	66	33	2	PTV-HD Rectum Bladder Femural head Penis Bulb Small bowel	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% D_{max}<69Gy, D_{mean} <30Gy, D_{50Gy}<33Gy D_{max}<65Gy, D_{mean}<30, D_{50Gy}<33Gy D_{max}<40Gy as low as possible D_{max}<50Gy	Pinnacle
I	Prostate gland seminal vescicles	54-72	30	1,8-2,4	PTV-HD PTV-LD	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% $V_{95\%}$≥95-98%	RayStation
	Prostate gland	67,5-74	30-37	2,25-2	Rectum Bladder Femural head Penis Bulb	V_{30Gy}<40% V_{50Gy}<20-25% V_{30Gy}<40% V_{50Gy}<20-25% D_{max}<34-36 Gy D_{max}<45Gy	
L	Seminal vescicles	60,75	27	2,25	PTV-HD PTV-LD Rectum Bladder	$V_{95\%}$≥95-98%; $V_{107\%}$≤2% $V_{95\%}$≥95-98% D_{max}≤68 - 70Gy; V_{65Gy}<15%; V_{40Gy}<40% V_{70Gy}<15%; V_{65Gy}<30% ; V_{40Gy}<60%	RayStation
	Prostate gland	70,2		2,7	Femural head Penis bulb	D_{max}≤ 40 - 45Gy D_{mean}<50 Gy; V_{40Gy}<50%;V_{60Gy}<25%	

					Small bowel	D_{max}≤45 - 48 Gy	

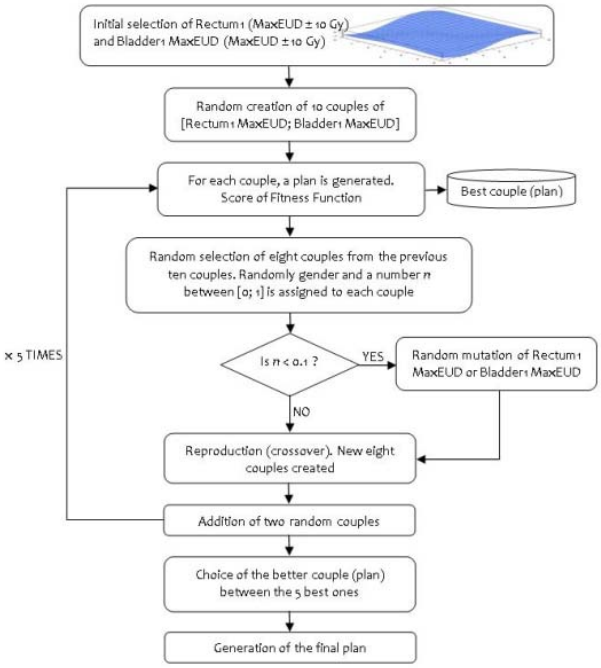
Table E1. Summary of dose prescription and constraints used for target and OARs for each centre.

	Parameters	Institutes	A	B	C	D	E	F	G	H	I	L
PTV	PTV_HD V_{95}	Manual	96.2 [94.7 97.4]	97.78 [96.0 99.0]	98.9 [98.0 100]	97.2 [95.1 99.2]	96.2 [94.0 100]	97.4 [96.1 99.0]	98.5 [97.12 99.89]	99.9 [99.2 100]	97.1 [94.6 100]	99.9 [99.6 100]
		Manual – AutoGPS	-2.63 [-4.18 -1.46] p<0.05 **A***	-0.9 [-3 1] p=0.62	2.1 [0 3.1] p=0.26	-0.13 [-2.13 3.72] 0.81	-0.08 [-3 2] p=0.079	-0.7 [-2.1 1.2] p=0.09	-0.51 [0.26 1.01] p=0.07	0.37 [-0.6 1.3] p=0.08	1.8 [-3.7 4.7] p=0.26	1.5 [0 3.94] p<0.05 **M***
	PTV_HD V_{107}	Manual	1.4 [0.1 4.8]	0.1 [0 0.1]	0.2 [0 1.6]	0 [0 0.1]	0.2 [0 0.9]	1.2 [0 4.9]	0 [0 0]	0 [0 0.3]	0.6 [0 5.4]	0 [0 0]
		Manual – AutoGPS	1.2 [-0.6 4.8] p<0.05 **A***	0.1 [0 0.8] p=0.34	0 [-1.5 1.6] p=0.94	0 [0 0.1] p=0.19	0.1 [-0.7 0.8] p=0.63	-1.5 [-6.4 0.2] p=0.05	0 [0 0] p= nv	-24.7 [-46.1 -7.2] p<0.001 **M***	0.31 [-1.1 4.7] p=0.54	0 [0 0] p= nv
	PTV_LD V_{95}	Manual	98.78 [93.09 100]	99.8 [99.3 100]	98.1 [96.1 99.4]	98.24 [94.1 100]		96.4 [89 100]			97.8 [96.6 99.7]	99.9 [99.3 100]
		Manual – AutoGPS	-1.08 [-6.55 0] p=0.07	-0.4 [-1.1 1.2] p= nv	1.89 [-1.1 5.9] p=0.86	2.6 [-0.2 7.6] p=n.v		-1.1 [-8.1 6.2] p=0.45			0.9 [-2.1 3.7]	1.2 [0. 4.3] p=0.71
	PTV_LD $CI_{50\%}$	Manual	5.91 [4.1 14.2]	4.54 [3.1 5.8]	26.44 [11.1 47.2]	15.0 [3.1 36.9]	3.3 [3 4]	3.52 [2.1 5.3]	3.75 [3.53 4.31]	8.31 [5.6 10.9]	7.2 [6.0 9.6]	36.89 [6.3 55.1]
		Manual - AutoGPS	0.65 [-1.08 4.34] p=0.06	0.15 [-1 1.09] p=0.61	2.44 [-1.1 7.8] p=0.06	2.4 [-0.4 7.6] p<0.05 **A***	-0.3 [-1 1] p=0.15	0.05 [-1.2 2.1] p=0.87	0.51 [0.26 1.01] p=0.06	1.5 [-2.5 5.4] p<0.05 **A***	1.3 [0.6 1.5]	9.25 [1.3 22.4] p=0.06
Rectum	Mean dose (Gy)	Manual	30.59 [26.56 42.18]	26.44 [16.8 32.7]	30.1 [20.2 36.2]	39.4 [27.1 63.3]	10.6 [7 15]	31.2 [24.1 41.2]	22.46 [19.75 25.53]	21.8 [14.1 26.9]	27.5 [17.6 31.9]	30.75 [28.1 35.2]
		Manual - AutoGPS	1.27 [-6.3 6.87] p=0.05	5.9 [-1 12] 0.03 **A***	7.7 [3.9 10.8] p<0.05 **A***	9.7 [3.8 30.9] p<0.05 **A***	-0.7 [-4 3] p=0.10	3.9 [-0.1 13] p<0.05 **A***	-2.72 [-6.97 1.76] p=0.03 **M***	0.4 -10.0 7.51] p=0.53	10.8 [5.8 15.9] p<0.05 **A***	10.4 [-0.9 24.5] p<0.05 **A***
	D[15.0%] (Gy)	Manual	55.46 [47.8 63.6]	49.22 [35.2 56.7]	54.8 [36.1 62.8]	61.7 [43.5 78.5	20.6 [11 25]	52.6 [44.1 62.3]	52.09 [41.08 57.48]	38.7 [7.7 57.1]	52 [34.5 57]	51.9 [46 60.7]
		Manual - AutoGPS	-1.84 [-7.1 9.8] p=0.06	4.7 [12.8 - 10.9] 0.28	7.6 [1.1 18.4] p=0.12	7.4 [0.7 22.2] p<0.05 **A***	-2.2 [-10 6] p=0.03 **M***	3.5 [-3.1 12.3] p=0.07	-1.12 [-4.81 4.74] p=0.10	6.1 [-43.8 34.2] p=0.81	21.2 [6.84 39.28] p<0.05 **A***	12.4 [8.1 17.3] p<0.05 **A***
Bladder	Mean dose (Gy)	Manual	22.79 [8.3 47.7]	30.5 [5.8 49.2]	23.8 [14.1 40.6]	43.5 [20.5 67.4]	9.2 [5 15]	37.2 [27.8 53.1]	19.94 [10.94 27.8]	45.7 [28.7 55.4]	28.8 [12.9 44.5]	31.4 [28.1 35.2]
		Manual - AutoGPS	2.06 [-4.8 6.5] p=0.05	20.9 [9.8 32.9] p=0.28	2.1 [-2.1 4.9] p=0.33	5.3 [-1.4 15.0] p<0.05 **A***	0.6 [-1 3] p=0.05	3.4 [-3.1 21.8] p=0.23	-0.37 [-3.99 1.61] p=0.09	5.4 [-18.7 16.3] p=0.39	3.2 [-1.47 14.29] p=0.26	12.4 [8.1 17.3] p<0.05 **A***
	D[25.0%] (Gy)	Manual	34.2 [5.5 67.9]	30.6 [5.8 48.7]	37.6 [17.2 65.9]	46.4 [11.8 69.4]	14.1 [4 24]	54.2 [45.1 66.7]	30.50 [9.65 49.71]	15.5 [9.3 19.7]	41.6 [15.3 67]	45.83 [29.6 60.7]
		Manual - AutoGPS	4.43 [-10.7 5.9] p=0.76	0.6 [-9.2 9.1] 0.29	3.6 [-5.1 10] p=0.33	4.7 [-2.7 16.7] p<0.05 **A***	1.2 [-2 6] p=0.99	1.4 [-1.8 9.9] p=0.3	-1.95 [-8.63 0.99] p=0.13	3.4 [-4.0 9.8] p=0.12	3.8 [-8.9 19.86] p=0.22	18.9 [11.4 26.5] p<0.05 **A***
Femural Head Left	Mean dose (Gy)	Manual	14.47 [9.9 19.5]	14.9 [9.8 18.2]	14.33 [12.1 16.8]	14.6 [10.5 18.4]	6 [3.1 8.2]	17.1 [12.9 21.8]	16.03 [7.62 19.98]	16.2 [8.9 22.2]	13.75 [6.2 18.9]	21.73 [17.6 27.9]
		Manual - AutoGPS	-0.76 [-5.0 4.8] p=0.44	0.7 [-2.9 4.8] 0.38	-1.3 [-5.2 0.9] p=0.08	1.6 [-1.1 3.45] p<0.05 **A***	-0.1 [-2.1 1.1] p=0.46	0.2 [-3.8 6.1] p=0.88	2.49 [-1.301 4.77] p<0.001 **A***	3.8 [-4.6 9.6] p=0.83	0.99 [-3.2 5.33] p=0.23	5.1 [1.4 10.9] p<0.001 **A***
Femural Head Right	Mean dose (Gy)	Manual	15.20 [10.4 20.4]	15.8 [8.9 19.7]	14.4 [12.8 16.9]	20.6 [10.2 47.8]	5.4 [3 8]	17.6 [12.8 23.6]	15.82 [9.55 21.94]	16.4 [[4.17 45.1]	13.7 [7.4 17.9]	20.17 [13.8 26.1]
		Manual - AutoGPS	0.33 [-3.9 7.0] p=0.11	1.9 [-3.1 4.8] 0.47	-1.2 [-3.9 0.8] p=0.24	4.9 [-1.4 17.1] p<0.05 **A***	-0.5 [-2 0] p=0.42	0.2 [-4.8 6.9] p=0.86	2.49 [-1.19 6.25] 0.07	1.4 [-7.5 11.7] p=0.3	0.58 [-3.4 3.6] p=0.38	3.5 [-0.2 8.49] p00.27

Penis Bulb	Mean dose (Gy)	Manual		15.1 [7.3 35.8]	33.3 [23.2 47.1]	44.35 [40.9 47.8]	8.1 [2 22]	33.2 [4.2 70]	53.31 [18.86 62.45]	0.68 [0.54 0.76]		25.9 [23.4 30.6]
		Manual - AutoGPS		-2.2 [-5.1 -3] 0.043 **M***	-5.2 [-9.1 1.1] p=0.69	16.7 [16.27 17.1] p=nv	-0.8 [-3 3] p=0.14	2.6 [-0.4 7.1] p<0.05 **A***	2.29 [-9.39 8.04] p=0.19	-0.03 [-0.2 0.1] p=0.24		1.6 [-13.47 3.04] p=0.85
	CBS	Manual - AutoGPS	10% [-15 20] p=0.13	0% [-20 20] p=0.56	7.5% [-10 15] p=0.375	19.5% [-40 40] p=0.19	-5% [-15 10] p=0.19	5% [-15 20] p=0.431	5% [-7 10] p<0.05 **A***	-31% [-43 -2] p<0.05 **M***	55% [-10 70] p<0.05 **A***	31% [8 44] p<0.05 **A***
	MU	Manual	892.9 [686 1194]	615.8 [463 859]	599.4 [546 729]		2664.7 [1889 3410]	746.2 [534 991]	957.1 [809 112]	582.7 [502 656]	569.1 [438 714]	602.4 [524 662]
		Manual - AutoGPS	-7.5 [-162 199] p=0.85	-160.8 [-304 -21] p<0.001 **M***	-239.3 [-444 0] p<0.001 **M***		44 [-1000 946] p=0.81	-63.6 [-368 90] p=0.22	197.1 [-99 365] p<0.05	-34.8 [-136 78] p=0.07	-102.1 [-231 8] p<0.001 **M***	-227.1 [-394 -39] p<0.001 **M***

Table E2. Summary of obtained plan parameter values in participating centres A to L. Population median values and ranges are reported for Manual. Differences with AutoGPS are characterized by population median values (upper), ranges (middle) and p-values (lower).

Paragraph E1



The above figure shows the workflow of the genetic algorithm; the chromosomes of the algorithm were the max equivalent uniform dose (MaxEUD) functions, which were applied to the ROIs created by subtraction between rectum and PTV (Rectum1) and between bladder and PTV (Bladder1).
An initial value of MaxEUD was selected by a 3rd degree polynomial function of two variables (volume of ROI and percentage of overlapping volume between original ROI and PTV). The function coefficients were derived from a baseline dataset of 50 patients already planned at the University of Turin radiotherapy clinic. Starting from these values, the range of MaxEUD was created by adding/subtracting 10 Gy ([MaxEUD±10 Gy]) (initial range). Ten couples of Rectum1 and Bladder1 MaxEUD were generated randomly ([Rectum1 MaxEUD; Bladder1 MaxEUD]). For each couple, a plan was generated with 20 optimization iterations and scored by the Fitness Function; coverage of the PTV was assured by a series of optimization parameters with the highest weight inside the objective function. The fitness function was defined as

$$FF = \sqrt{\frac{2(CI)^2 + 4\left(\frac{rAD}{PD}\right)^2 + 2\left(\frac{bAD}{PD}\right)^2 + \left(\frac{lfAD}{PD}\right)^2 + \left(\frac{rfAD}{PD}\right)^2}{10}}$$

where CI and PD are respectively the conformity index (calculated as the ratio between the ROI volume covered by the isodose and the total isodose volume) and the prescribed dose to PTV while rAD, bAD, lfAD and rfAD are the average dose of rectum, bladder and femoral heads (left and right) respectively. The plan with the best FF (the lowest value) was saved and archived. Eight couples were then randomly selected from these initial ten; a randomly selected gender (female or male) and a number (n) between [0; 1] were then assigned to each of them. If n is inferior to 0.1 (mutation probability value), one of the two elements of the couple was randomly changed within the initial range established. The eight couples evolved by crossover action and new eight couples were created. Two new random couples were added to the eight (making a total of ten) and ten more plans were generated from these new couples. This workflow was repeated four times (for a total of 5 cycles). At the end, five plans were selected from the best couples ([Rectum1 MaxEUD; Bladder1 MaxEUD]). The best plan in terms of FF was selected and the final plan was calculated (0.3 cm dose grid and 40 optimization iterations).