Quality assessment for VMAT prostate radiotherapy planning based on data envelopment analysis

Kuan-Min Lin[1], John Simpson[1,2], Giuseppe Sasso[3,4], Andrea Raith[1] and Matthias Ehrgott[5,1]

[1] Department of Engineering Science, University of Auckland, 70 Symonds Street, Auckland, New Zealand

[2] Auckland Radiation Oncology, 98 Mountain Road, Epsom, Auckland, New Zealand

[3] Auckland City Hospital, 2 Park Road, Grafton, Auckland, New Zealand

[4] Discipline of Oncology, School of Medical Sciences, University of Auckland, 85 Park Road, Grafton, Auckland, New Zealand

[5] Department of Management Science, Lancaster University, Lancaster LA1 4YX, United Kingdom

Email: km.lin@auckland.ac.nz

Abstract

The majority of commercial radiotherapy treatment planning systems requires planners to iteratively adjust the plan parameters in order to find a satisfactory plan. This iterative trial-and-error nature of radiotherapy treatment planning results in an inefficient planning process and in order to reduce such inefficiency, plans can be accepted without achieving the best attainable quality. We propose a quality assessment method based on Data Envelopment Analysis (DEA) to address this inefficiency. This method compares a plan of interest to a set of past delivered plans and searches for evidence of potential further improvement. With the assistance of DEA, planners will be able to make informed decisions on whether further planning is required and ensure that a plan is only accepted when the plan quality is close to the best attainable one. We apply the DEA method to 37 prostate plans using two assessment parameters: rectal generalized equivalent uniform dose (gEUD) as the input and D95 (the minimum dose that is received by 95% volume of a structure) of the planning target volume (PTV) as the output. The percentage volume of rectum overlapping PTV is used to account for anatomical variations between patients and is included in the model as a non-discretionary output variable. Five plans that are considered of lesser quality by DEA are re-optimized with the goal to further improve rectal sparing. After re-optimization, all five plans improve in rectal gEUD without clinically considerable deterioration of the PTV D95 value. For the 5 re-optimized plans, the rectal gEUD is reduced by an average of 1.84 Gray with only an average reduction of 0.07 Gray in PTV D95. The results demonstrate that DEA can correctly identify plans with potential improvements in terms of the chosen input and outputs.

Mathematics Subject Classification: 90C90, 90C29 and 90C05

1. Introduction

During the past two decades, radiotherapy treatment techniques have advanced significantly due to the development of dynamic multileaf collimators coupled with commercially available inverse planning systems. This has led to the widespread use of Intensity Modulated Radiation Therapy (IMRT) which facilitates improved conformance to the treatment volume while minimizing dose to surrounding avoidance structures. Traditional fixed-gantry IMRT has been further developed to arc-based IMRT, which includes Intensity Modulated Arc Therapy (IMAT) (Yu and Tang 2011) and tomotherapy (Mackie *et al* 1993). Since arc-based IMRT can utilize many more angles to deliver the radiation than fixed-gantry IMRT in a given treatment time, it provides better flexibility in shaping the dose distribution and thus has the potential to achieve a better treatment outcome (Shepard and Cao 2011). Combining these treatment modalities with image guidance (Xing *et al* 2006), the ability to deliver complex dose distributions accurately has become a reality. However, despite the advancement of these technologies, maximizing the benefits afforded by them can be a challenging task due to the difficulties associated with radiotherapy treatment planning.

Radiotherapy treatment planning involves managing several conflicting objectives related either to the planning target volume (PTV) or healthy organs at risk (OARs). The major commercial treatment planning systems use a 'weighted sum of the constituent objectives' technique in deriving radiotherapy treatment plans. Radiotherapy treatment planning in this approach involves a planner entering a number of plan objectives into the treatment planning system. Note that the chosen plan objectives may not be the same as those used to evaluate the plan acceptability, but rather are selected based on the personal experience of the planner, see Phillips and Holdsworth (2011) for a discussion of planning objectives versus decision objectives for the comparison of treatment plans. Each of these objectives is associated with an importance score (weight). The treatment planning system derives treatment plans based on the plan objectives and the associated weights. During the planning process, the planner iteratively adjusts the plan parameters (i.e. the objective weights and/or the objectives) to improve plan quality. However, because the exact effects of changing the plan parameters cannot be known a priori, it is hard for the planner to verify if there is further potential to improve a plan. If a plan is deemed to be of inadequate quality it would take further time to produce another plan, without knowing in advance whether the new plan is going to be superior to the previous plan. This trial-and-error aspect of the planning process is inefficient and in order to reduce such inefficiency, plans can be accepted without achieving the full potential of the available technology.

This planning dilemma can be addressed by comparing the plan quality against past plans. This comparison will allow the planners to score the plan against an "optimal" plan and therefore allow informed decisions on whether further improvement is possible. The "optimal" plan may not be truly the best one achievable but rather an indication of what can be achieved. A number of plan assessment approaches that use past plans as references have been proposed in the literature. Hunt *et al* (2006) predict the achievable mean dose to the parotid gland (OAR) for head and neck IMRT plans using a linear model based on the percent volume of the parotid gland overlapping the PTV(s). Similarly, Moore *et al* (2011) build a nonlinear model to predict the achievable mean dose to the OAR based on the percentage volume of OAR overlapping the PTV. Their model is more general than that of Hunt *et al* (2006) as the predicted mean dose to the OAR is normalized by the prescription dose and the dataset used to build the model contains both parotid glands of head-and-neck IMRT plans and rectums of prostate plans. Wu *et al* (2009) use a case based approach in which the minimum achievable dose to the OAR is queried from past plans with the same or more challenging OAR-PTV geometric relationship. In their study, the OAR-PTV geometric relationship is described by a so-called overlap volume histogram (OVH), which measures the percentage volume of an OAR that is within a specified distance of a PTV. The use of OVH is based on the concept that OARs closer to the target are harder to spare while OARs further away from the target are easier to spare. Zhu *et al* (2011a) build a model that predicts the achievable dose volume histogram (DVH) of the OAR based on a given distance-to-target histogram (DTH), which is a histogram similar to OVH (Zhu *et al* 2011b). They assess the plan quality by comparing the OAR DVH to the predicted DVH. In the study, principal component analysis is used to identify the most significant components for the DTHs and the DVHs of the OARs. Using these components as the parameters, support vector regression is then used to build the model for prediction.

These quality assessment approaches predict the achievable OAR sparing based on the geometrical relationship of the PTV and the OAR. Thus, given a particular geometrical relationship, if a newly generated plan has a significantly higher dose to the OAR than the predicted dose, the plan is considered of inadequate quality and re-optimization is required. However, as these approaches do not consider the dose to the PTV, one may unintentionally conclude that more OAR sparing is available without realizing that the improvements in OAR sparing would likely deteriorate the PTV dose coverage. This might inadvertently lead to a longer planning process since a high quality plan with good PTV coverage and acceptable OAR sparing may be considered of inadequate quality simply because it does not achieve the maximal OAR sparing.

In this study we propose a plan quality assessment method that can take both the dose to the PTV and the OARs into consideration. By doing so, when a plan is generated, a planner can assess the plan quality using its current dose distribution to both the PTV and the OARs rather than assess it against a plan that solely encourages maximal OAR sparing. The method is based on Data Envelopment Analysis (DEA) (Charnes *et al* 1978, Cooper *et al* 2011). DEA is a management science method for assessing the performance of a set of decision-making units (DMUs) that convert inputs into outputs.

In a loose economic interpretation, the inputs represent the cost we pay for producing outputs. The concept of DEA is directly applicable to the problem of assessing treatment plan quality in radiotherapy in which the doses to OARs are considered as the cost we pay for delivering dose to the PTV. In fact, DEA performs peer evaluation by comparing treatment plans with respect to an ideal defined by historical plans. In healthcare, DEA has been applied in performance assessment of healthcare systems (Chilingerian and Sherman 2011), including formative evaluation of radiotherapy services (Santos and Amado 2012) and even to compare prostate cancer treatment options (Ramer *et al* 2008). However, to the best of our knowledge, it has never been used for case-based quality assessment for radiotherapy treatment planning. One of the most valuable strengths of DEA is its ability to handle multiple inputs and outputs. This strength makes DEA ideal for radiotherapy plan assessment in which several conflicting planning criteria need to be considered.

The purpose of this article is to demonstrate how DEA can be applied to assess the quality of radiotherapy treatment plans. Despite the capabilities of DEA, in this initial study of prostate treatment plans, we focus on a DEA model that considers a single input and a single output as well as an environmental variable. This paper is an extended version of Lin *et al* (2012) in which we described the mathematical aspects of the method. The current paper contains extended clinical results and discussion of the proposed method. In section 2, we introduce the mechanism of DEA. In section 3 we present a case study where DEA is applied to assess the quality of prostate radiotherapy plans. The results and discussion are presented in sections 4 and 5 respectively.

2. Introduction to DEA

In this section we introduce concepts of DEA before we demonstrate its application to radiotherapy treatment planning in section 3. For a detailed introduction to DEA we refer the interested reader to chapter 6 of Coelli *et al* (2005) and to Cooper *et al* (2011).

In DEA, decision-making units (DMUs) are assessed based on their ability to convert a set of inputs into outputs. This ability is referred to as efficiency. Thus, given a set of DMUs, DEA assesses the efficiency of each DMU by comparing it with all other DMUs and searching for evidence of potential improvements, i.e. whether some of its inputs or outputs can be improved without worsening any of its other inputs or outputs. If there is no evidence for potential improvements, the DMU is considered fully efficient. In general, DEA can be conducted in an input orientation or an output orientation. In an input oriented model, DEA searches for potential improvement of a DMU through proportional reduction of the inputs. Alternatively, in an output oriented model, the potential improvement of a DMU is found by proportional augmentation of the outputs. In this study we focus on input oriented models. Formulations of output oriented models can be found in Cooper *et al* (2011). In the radiotherapy application decsribed in this paper, the DMUs will be treatment plans and the inputs will be dose delivered to OARs, which can be considered to be cost for delivering dose to the PTV (the output).

An intuitive way to explain how DEA operates is through the envelopment model. Assume there are I DMUs each with N inputs and M outputs. For the ith DMU the inputs and the outputs are represented by vectors $x^i \in \mathbb{R}^N$ and $q^i \in \mathbb{R}^M$, respectively. The data for all I DMUs can be represented by the input matrix $X \in \mathbb{R}^{N\times I}$ and the output matrix $Q \in \mathbb{R}^{M\times I}$ in which the ith column contains the data for the ith DMU. The efficiency score θ^i for the ith DMU is derived by solving the following DEA linear programming model in envelopment form:

$$\begin{aligned} \min\ & \theta^i \\ s.t.\ & -q^i + Q\lambda \geq 0 \\ & \theta^i x^i - X\lambda \geq 0 \\ & e^T\lambda = 1 \\ & \lambda \geq 0, \end{aligned} \qquad (1)$$

where $\theta^i \in \mathbb{R}$ and $\lambda \in \mathbb{R}^I$ are the decision variables and $e \in \mathbb{R}^I$ is a vector of ones. Note that this optimization problem needs to be solved I times, once for each of the I DMUs, if one wishes to derive the efficiency score for all of the DMUs. The optimal solution of (1) is θ^{i*} and λ^{i*}, where θ^{i*} is the efficiency score of the ith DMU and λ^{i*} is a vector of weights. The formulation attempts to maximally

scale down the input vector x^i while maintaining the scaled input vector θx^i within the feasible set, which is defined by all convex combinations of the existing DMUs, i.e. $\{(Q\lambda, X\lambda): e^T\lambda = 1, \lambda \geq 0\}$. The feasible set is a set of potentially attainable DMUs constructed based on inputs and outputs of the existing DMUs. DEA attempts to search for evidence of further improvements of the inputs among these potential DMUs, i.e. in our application it looks for evidence that reduction in OAR dose is possible without decreasing PTV dose. In the circumstance where no scaling is possible, i.e. when $\theta^{i*} = 1$, the ith DMU is identified as fully efficient.

Conceptually, scaling the input vector x^i shifts the DMU i to a projected point $(Q\lambda^{i*}, X\lambda^{i*})$ on the efficient frontier, which is a multidimensional surface formed by all of the points representing efficient DMUs. The projected point $(Q\lambda^{i*}, X\lambda^{i*})$ is also referred to as the target for the ith DMU. The target is formed by a convex combination of existing DMUs with non-negative weights λ^{i*} summing to 1. The existing DMUs which contribute to the target are referred to as the peers of the ith DMU. The target represents the inputs and outputs that the ith DMU should aim for to make itself efficient. In treatment plan assessment, the target indicates the extent to which other treatment plans suggest further OAR sparing can be achieved. For an efficient DMU, the DMU itself is its own target as well as its only peer. Note that, in practice, we do not need to establish the efficient frontier to perform DEA. Instead, we solve formulation (1) once for each DMU.

The mechanism of DEA can be illustrated using figure 1, which shows a set of DMUs (A, B, C, D, E) with one input and one output. The efficient frontier is shown by the piecewise linear solid line connecting efficient DMUs A, B, and C. These DMUs have a more preferable output to input ratio than the inefficient DMUs D and E, for the corresponding input level. DEA assesses the efficiency of DMU D and E by scaling the input required to shift these DMUs to their corresponding targets D* and E* on the efficient frontier. DMUs A, B and C are considered fully efficient since no scaling of the input is required to shift these DMUs to the frontier.

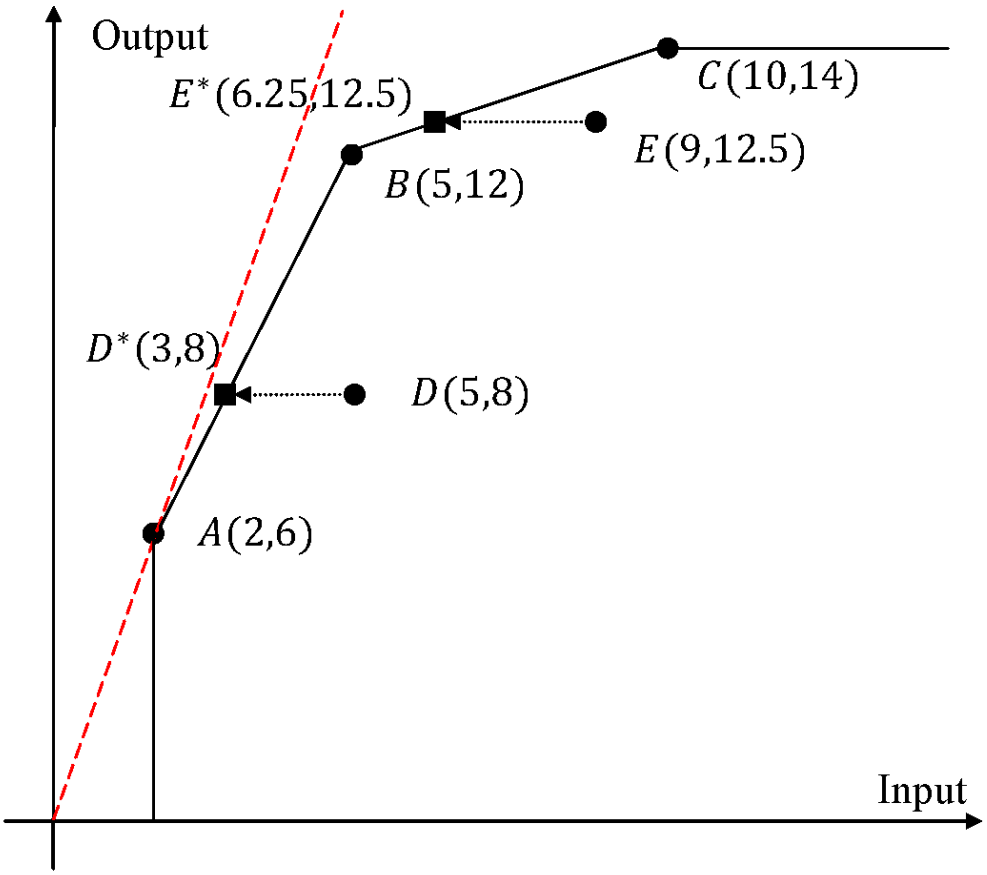

Figure 1. The efficient frontier of a CRS model (dashed line) and a VRS model (solid line).

The efficiency scores, input and output values, target values and the corresponding peers and weights of the DMUs in figure 1 are summarized in table 1. It is shown clearly that target point D* is formed by convex combination of DMUs A and B with weights 0.6667 and 0.3333, respectively. Similarly, target point E* is formed by convex combination of DMUs B and C with weights 0.75 and 0.25, respectively.

Table 1. The efficiency score, input, output, target, corresponding peers and weights for the five DMUs shown in figure 1.

DMU	Efficiency score	(Input, Output)	Target (input, output)	Peer(s)	Weight
A	1.000	(2,6)	(2,6)	A	1.000
B	1.000	(5,12)	(5,12)	B	1.000
C	1.000	(10,14)	(10,14)	C	1.000
D	0.600	(5,8)	(3,8)	A	0.667
				B	0.333
E	0.694	(9,12.5)	(6.25,12.5)	B	0.750
				C	0.250

The model in formulation (1) assumes variable returns to scale (VRS), which suggests that as the inputs change the outputs change at a variable proportion and vice versa. Essentially, the VRS model restricts the feasible set to the convex combinations of the existing DMUs, the efficient frontier is therefore a piecewise linear function that connects the efficient DMUs, e.g. DMUs A, B and C in figure 2.

Alternatively we could use the assumption of constant returns to scale (CRS) by removing the constraint $e^T\lambda = 1$, from formulation (1). CRS suggests that the return of outputs for each unit of input is the same for all input levels. In the CRS model, the feasible set is given by a line (hyperplane) through the origin with the shallowest gradient that has all DMUs to its right. The efficient frontier is shown by the red dotted line in figure 1.

When performing DEA, one may want to take environmental factors into account. Environmental factors are factors that may influence the performance of DMUs but are external and out of the control of the DMUs. In our study, the percentage overlap of PTV and rectum is such an environmental variable – this value is clearly not influenced by the treatment plan but can adversely affect the attainable quality of the treatment plan. Environmental factors can be incorporated in the DEA formulation as environmental variables. A number of ways to handle environmental variables are introduced in chapter 7 of Coelli *et al* (2005). In this study, we assume that higher values of the environmental variables are likely to impair the efficiency of the DMUs and we incorporate the environmental variables into the DEA model as a so-called non-discretionary output variable (Banker and Morey 1986). We assume there are L environmental variables represented by a vector $z^i \in \mathbb{R}^L$ for the ith DMU and by a matrix $Z \in \mathbb{R}^{L\times I}$ for the whole dataset. Environmental variables are incorporated into the DEA formulation by adding the constraint $-z^i + Z\lambda \geq 0$ into formulation (1). The constraint restricts the feasible set and ensures that DEA searches for the evidence of further improvements among DMUs that are associated with higher or equal values of environmental variables than the ith DMU. Since high values of the environmental variables are considered unfavourable to the efficiency, if a DMU can achieve certain efficiency with higher or equal values of environmental variables than the ith DMU, the ith DMU should be able to achieve at least the same efficiency, otherwise it is considered inefficient. For completeness, the DEA formulation used in this study, which is an input oriented VRS model with a non-discretionary environmental output variable, is shown in formulation (2).

$$\begin{aligned} &\min\ \theta^i \\ s.t.\ \ &-q^i + Q\lambda \geq 0 \\ &\theta^i x^i - X\lambda \geq 0 \\ &-z^i + Z\lambda \geq 0 \\ &e^T\lambda = 1 \\ &\lambda \geq 0. \end{aligned} \tag{2}$$

3. Application of DEA to Prostate Radiotherapy Treatment Plans

One of the most difficult tasks in prostate radiotherapy treatment planning is managing the dose delivered to the PTV and the rectum. The rectum is usually the OAR that most influences the ability to achieve the optimal dose to the PTV. In this preliminary study, we only consider the dose delivered to the PTV and the rectum. The goal is to generate a dose distribution that matches the prescription dose in the PTV as closely as possible while maximizing rectal sparing. In a loose economic interpretation, the dose delivered to the rectum is considered as the cost for delivering the dose to the PTV. Specifically, we use D95 of the PTV as the output and generalized equivalent uniform dose (gEUD, Niemierko 1999) of the rectum as the input for the DEA model. In addition, we use the percentage volume of the rectum that overlaps the PTV as an environmental variable. The higher the overlap the more difficult it is to achieve good PTV coverage and low OAR dose simultaneously. We handle the environmental variable using the method described in section 2. While many other dose descriptors or anatomical descriptors can be alternatively used for the DEA model, it is out of the scope of this study to investigate the most preferable inputs and outputs. Instead, we empirically select these input and outputs and focus on investigating the validity of using DEA as a quality assessment method in radiotherapy planning.

We use an input oriented model for the analysis since we are interested in maximal OAR sparing available for a given dose to the PTV. Mathematically, we may assume CRS for the model since we are able to obtain a constant return of PTV D95 for each unit change of rectum gEUD by scaling the entire dose distribution. However, because the treatment plans used in our case study all have a small variation in PTV D95 (standard deviation 0.3 Gy) versus a standard deviation of 1.17 Gy in rectal gEUD, the CRS approach is not appropriate in our context. Since the DMUs with their different input levels are all clinically acceptable plans, a VRS model offers a better approximation of what is clinically attainable for a given input level than a CRS model.

A series of 37 anonymized clinically intact prostate treatment plans were provided by Auckland Radiation Oncology, following approval and guidelines of New Zealand Health & Disabilities Ethics Committees for observational study. All plans were the actual plans used for the subsequent delivery of treatment and were generated using the same treatment planning system over a 1 year period utilising the same plan acceptability criteria. In all cases, 74 Gray (Gy) was prescribed to the PTV with requirements that 99% of the PTV received 95% of the prescribed dose and that 99% of the actual prostate receive 99% of the prescribed dose. For rectal criteria, the percentage volume of the rectum that received equal to or more than 40Gy, 60Gy and 70Gy of radiation dose should not exceed 60%, 40% and 10%, respectively i.e. V40Gy<60%, V60Gy<40% and V70Gy<10%. Not all rectal criteria were met in all cases, but all plans were nevertheless considered clinically acceptable. All plans were planned for VMAT delivery with Pinnacle v9 and the SmartArc module (Philips, Netherlands) using a single 360 degree arc. The plan input/output values were extracted using CERR (Deasy *et al* 2003). The input/output values of the plans are shown in table 2. We used an in-house DEA software package, pyDEA (Raith *et al* 2012), to assess the efficiency of these 37 prostate plans. After obtaining the results from the analysis, five inefficient plans were selected for re-optimization. We use the term re-optimization throughout this paper to describe the process of modifying the treatment planning objectives and re-running the inverse plan optimizer. The input/output values of the re-optimized plans are also included in table 2, indicated by the original plan ID with an asterisk. Each of the selected plans has a percentage overlap volume substantially different from the other selected plans. These plans are selected in order to test the ability of DEA in assessing plans with variations in anatomical structure relationships. A planner was instructed to further improve rectal sparing while maintaining overall clinical acceptance for the selected plans without access to the results of DEA. The original plans were optimized for rectal sparing via the use of a single rectal objective based on gEUD with the volume parameter "a" equal to 1. Plan re-optimization involved changing the a-value from 1 to 6 thereby increasing the penalty weight of the higher dose components of the rectal DVH. All other aspects of the plan remained the same. After re-optimization, the plans were included in the dataset and the DEA analysis repeated.

Table 2. The efficiency scores and the input/output values of the plans. Plans 10, 19, 26, 31 and 35 were re-optimized. Those plan IDs with an asterisk indicate re-optimized plans.

Plan ID	Efficiency (original dataset)[a]	Efficiency (re-optimized dataset)[b]	Output: D95 PTV (Gy)	Input: Rectal gEUD (Gy)	Fractional overlap
1	0.988	0.975	71.625	61.418	0.048
2	0.982	0.980	71.025	62.478	0.100
3	0.966	0.960	71.275	62.805	0.066
4	0.985	0.977	71.225	61.369	0.053
5	0.994	0.981	71.675	60.744	0.036
6	0.974	0.968	71.825	63.714	0.077
7	0.982	0.975	71.675	62.779	0.074
8	0.978	0.968	71.775	61.745	0.031
9	0.982	0.972	71.775	62.560	0.061
10	0.964	0.955	71.575	63.457	0.065
10*	N/A	0.987	71.525	61.314	0.065
11	0.982	0.970	71.575	61.730	0.049
12	0.979	0.969	71.525	61.610	0.042
13	0.985	0.975	71.675	62.160	0.062
14	0.998	0.987	71.725	60.362	0.030
15	0.980	0.966	71.625	61.934	0.047
16	0.984	0.976	71.475	62.192	0.072
17	0.995	0.986	71.475	61.403	0.069
18	1.000	0.997	71.825	60.173	0.022
19	0.975	0.963	71.725	62.562	0.052
19*	N/A	0.987	71.325	60.710	0.052
20	0.994	0.985	71.375	60.502	0.037
21	1.000	1.000	71.975	63.648	0.111
22	0.978	0.977	72.025	63.870	0.076
23	0.985	0.982	71.875	63.644	0.091
24	0.983	0.983	72.075	63.763	0.077
25	1.000	1.000	72.125	61.192	0.033
26	0.980	0.980	71.325	63.545	0.110
26*	N/A	0.991	71.025	62.181	0.110
27	0.976	0.969	71.425	62.746	0.075
28	1.000	1.000	72.075	62.019	0.062
29	1.000	1.000	70.875	61.572	0.115
30	0.986	0.980	71.625	62.603	0.079
31	0.978	0.968	71.425	61.522	0.038
31*	N/A	1.000	71.725	59.567	0.038
32	1.000	1.000	72.125	64.804	0.119
33	0.993	0.990	72.025	61.854	0.046
34	0.990	0.978	71.775	61.593	0.049
35	0.966	0.960	71.425	63.570	0.080
35*	N/A	0.993	71.525	61.658	0.080
36	0.974	0.970	71.175	62.529	0.079
37	1.000	0.992	71.575	60.056	0.032
Average	0.985	0.980	71.611	62.098	0.064

Standard deviation	0.010	0.012	0.302	1.170	0.025

[a] Efficiency score according to DEA based on original treatment plans 1 to 37 (original dataset).

[b] Efficiency score according to DEA based on original treatment plans 1 to 37 and re-optimized plans 10*, 19*, 26*, 31* and 35* (re-optimized dataset).

The data of table 2 is visualized in figure 2. Notice that due to the presence of the environmental variable, the efficient frontier is not a single piecewise linear line as in figure 1.

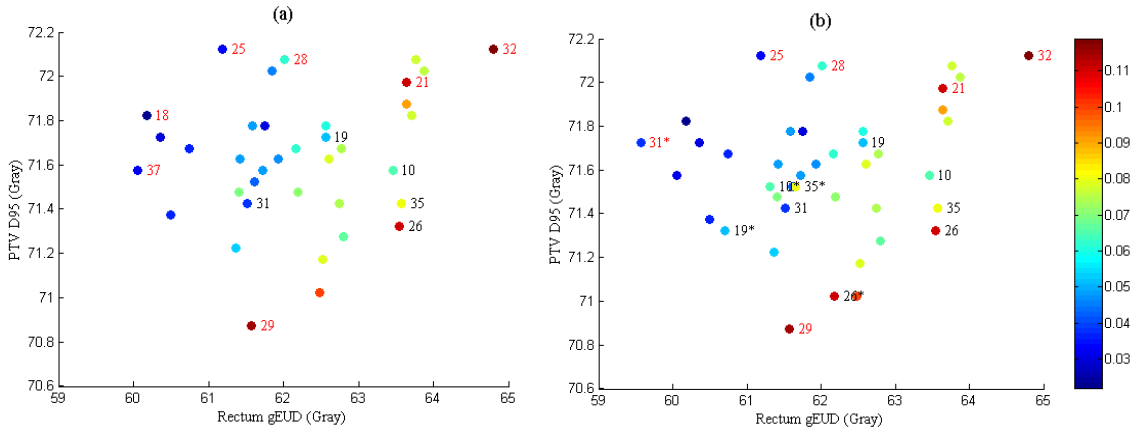

Figure 2. Plot of the data points where colour represents PTV rectum overlap, (a) before and (b) after re-optimization of a subset of the plans. Red numbers indicate treatment plans considered efficient in each DEA analysis. Black numbers indicate treatment plans selected for re-optimization.

4. Results

The efficiency scores before and after including re-optimized plans as well as the input and outputs used in DEA are provided in table 2. The high efficiency scores (average 0.985, standard deviation 0.01) indicate that the plan qualities of these prostate plans, in terms of PTV D95 and rectal gEUD, are quite consistent in general. The PTV D95 values among these plans vary slightly from 70.88 to 72.13 Gy while rectal gEUD and percentage overlap volume vary quite considerably from 59.57 to 64.80 Gy and from 2.2% to 11.9%, respectively. We find very strong evidence from the original dataset that percentage overlap volume is correlated to rectal gEUD ($p = 1.76 \times 10^{-8}$). This suggests that one reason for the large variations in rectal gEUD is due to the large variations in percentage overlap volume. The small variation of PTV D95 among the plans is consistent with the fact that meeting the PTV coverage requirement was the highest priority for the planners unless instructed otherwise.

Plans 10, 19, 26, 31 and 35 were identified as substantially inefficient for their range of percentage overlap volume and were re-optimized. The efficiency scores of these plans are less than the original average efficiency score (0.985). The re-optimization of plan 31 produced an additional efficient plan in the dataset. This plan extends the efficient frontier slightly and results in lower or equal efficiency scores of all other plans compared to those of the original dataset, as shown in table 2. The efficiency score of the re-optimized plans are higher than those of the original plans, with an average improvement of 0.026 units. Note that this improvement is quite substantial since the standard deviation of the efficiency scores is only 0.012 (table 2). All of the re-optimized plans achieved an efficiency score higher than the re-optimized average efficiency score (0.98).

A clinical peer review of the plans was performed using MIM Maestro V5.4 Mac Version (MIM Software Inc. Cleveland, OH). In this review, clinical and DVH parameters not included in the DEA (ie. bladder, femoral heads, dose maxima, hot spot percentage and site, etc) were also taken into account. The review confirmed that re-optimized plans 10*, 31* and 35* were deemed superior when comparing with the original plans. Figure 3 shows improved conformity for these re-optimized plans. The green shaded areas show iso-dose volume of the original plans and the red shaded areas show the improvement in iso-dose volume after re-optimization. Re-optimized plan 19* was deemed substantially equivalent to the original plan from a clinical and dosimetric viewpoint. For plan 26, the

re-optimized plan was not considered superior to the original plan due to comparatively inferior dose painting and conformity. However, we note that dose painting and conformity index were not taken into account in the DEA analysis and the goal of re-optimization was to further improve rectal sparing while maintaining overall plan acceptance, which was achieved for all five re-optimized plans. The fact that planners could be instructed to achieve better rectal sparing while maintaining overall plan acceptance and subsequently were able to do so, is a positive finding.

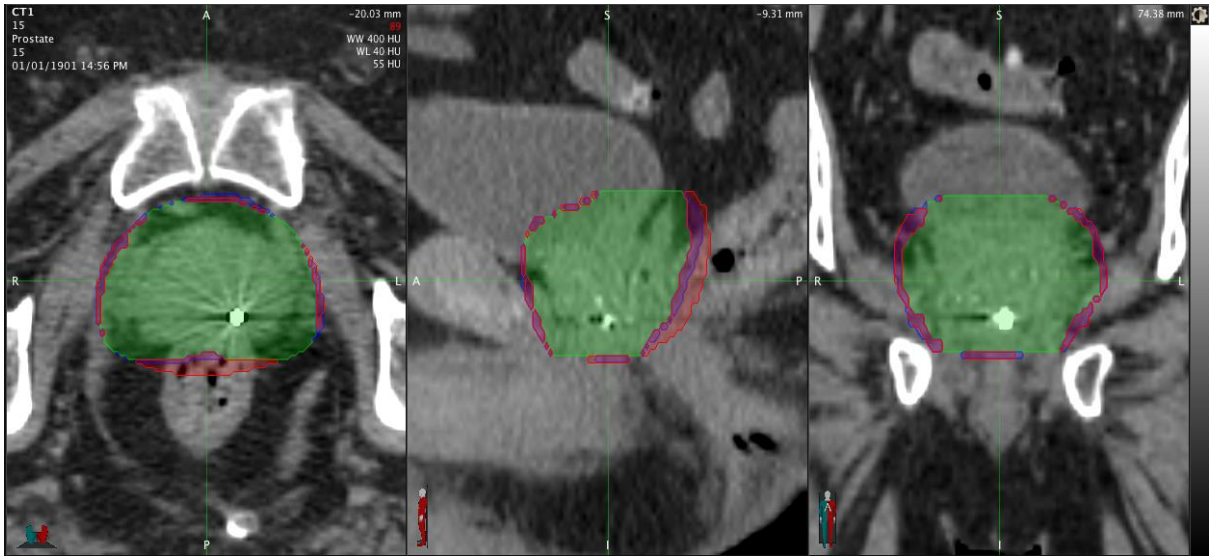

Figure 3. Examples of improved conformity for re-optimized plans.

Table 3. The original, re-optimized and the corresponding target input values for the selected plans. The measurement unit for the input is Gray.

Plan	Original	Target (original dataset)	Re-optimized	Target (re-optimized dataset)	Dose reduction[a]	Prediction error[b]
10	63.457	61.198	61.314	60.508	2.143	0.806
19	62.562	61.007	60.710	59.930	1.852	0.780
26	63.545	62.291	62.181	61.632	1.364	0.549
31	61.522	60.167	59.567	59.567	1.955	0.000
35	63.570	61.409	61.658	61.211	1.912	0.447
Avg	62.931	61.214	61.086	60.570	1.845	0.516

[a] Dose reduction is calculated as original input value minus re-optimized input value.

[b] Prediction error is the absolute difference between the re-optimized target input value and the re-optimized input value

The original, re-optimized and the target input values for the selected plans are summarized in table 3. The overlap fractions are not included since they are the same as the corresponding values in table 2. In terms of the chosen input and outputs, re-optimization of the five plans resulted in an average reduction of 1.84 Gy in rectal gEUD. The reduction in PTV D95 is only between -0.3 (an actual increase) and +0.4 Gy with an average of +0.07. Due to this minor change in PTV D95 only rectal gEUD (input) is shown in table 3. Among the five plans, the minimum rectal gEUD reduction is 1.364 Gy (table 3). These results suggest that DEA successfully identified plans with potential improvements in terms of the chosen input and outputs. In addition, although planners were not provided with the results of DEA, the values for the re-optimized plans are very close to their targets, with a maximum difference of 0.806 Gy. Note that the target of a re-optimized plan represents DEA's prediction of best attainable plan. This minor input/output difference between the prediction and the re-optimized plan demonstrates the ability of DEA in predicting potential improvement in terms of the input and outputs. The DVHs for the original plans and the re-optimized plans are shown in figure 4.

The improvements in rectal sparing are clearly illustrated in the DVHs while there are no clinically considerable differences between original PTV DVHs and re-optimized PTV DVHs.

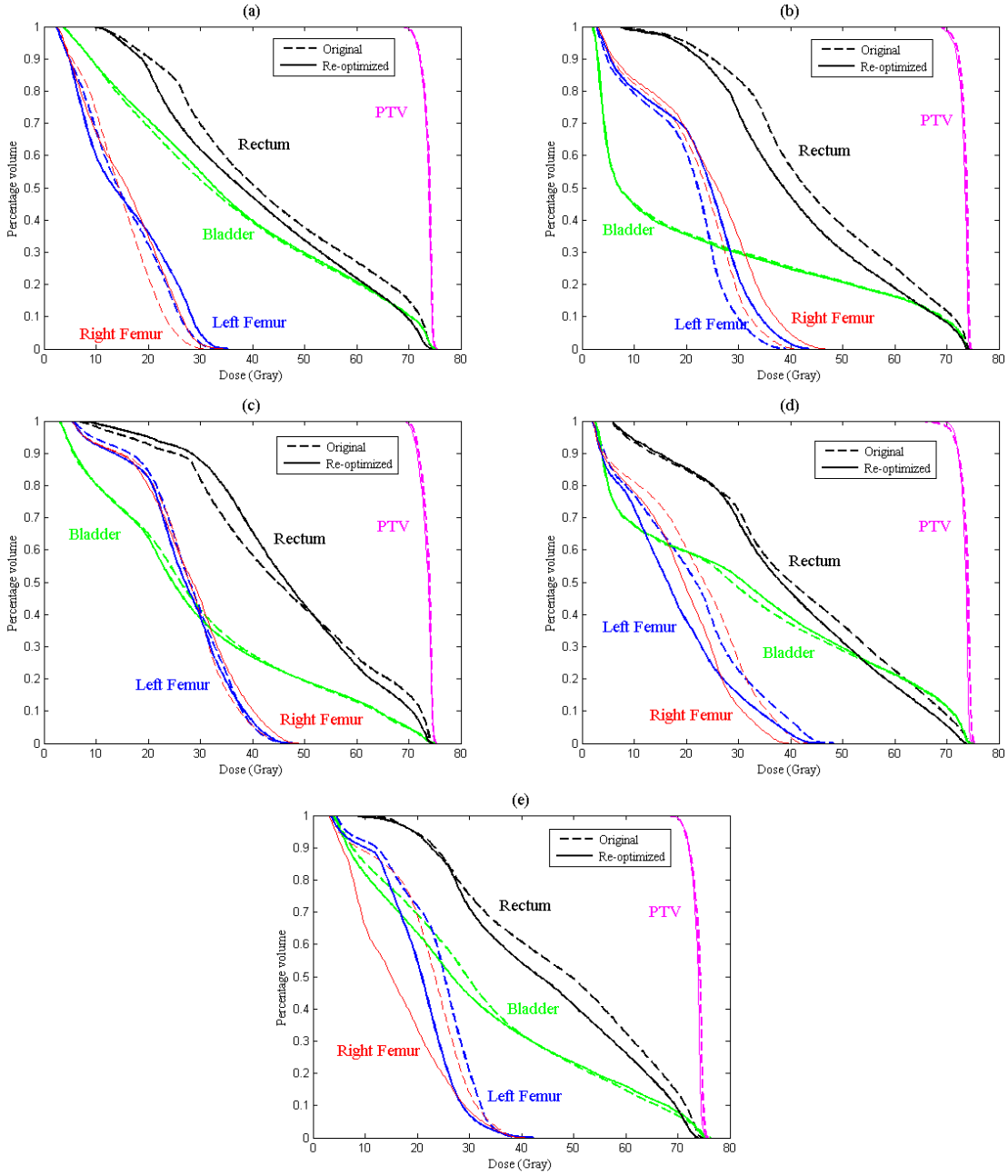

Figure 41. The original and re-optimized DVHs of (a) plan 10, (b) plan 19, (c) plan 26, (d) plan 31 and (e) plan 35.

To provide an assessment of the clinical relevance of the plan changes, an analysis of the biological objective function P+ was undertaken. P+ is a scalar quantity that combines the probability of tumour control with the probability of normal tissue complication (Kallman *et al* 1992) and is often referred to as the complication free tumour control probability. The average P+ value increased from 0.618 to 0.632, a 2.3% improvement. If, however, we scale the treatment plan to that of a 78Gy prescription, which is now our institutional norm for medium to high risk cases, then the average P+ increases from 0.661 to 0.685, a 3.6% improvement. As the dose prescription increases, the biological consequences and potential gains also increase. In deriving P+, the endpoint for normal tissue complication was necrosis/stenosis according to the values of Ågren-Cronqvist (1995) and tumour control probability from Cheung *et al* (2005).

An additional measure that can be extracted from DEA analysis is the peer count, which is the number of times a plan is referred to as a peer. The peer count indicates how often an efficient DMU is used as benchmark for others and thus attests to the plan's quality. The peer counts of the efficient plans are shown in table 4. There are 7 efficient plans in the original dataset. After including plan 31* in the dataset, plan 18 and plan 37 are no longer efficient and are not used as a peer for other plans. Plan 25 and plan 32 have relatively low peer counts for both the original and re-optimized dataset, as shown in table 4. One limitation of DEA is that a plan can be considered efficient simply because it has the best value in one of the inputs or outputs. Plan 25 has the highest PTV D95 value and plan 32 has the highest overlap fraction. However, the peer counts of these two plans being greater than 1 suggests that these plans have preferable output to input ratios compared to at least one other plan and that they are not considered efficient purely because they have an optimal value in one of the inputs or outputs. Plan 31* has the highest peer count of 37 and is referred to as a peer for all inefficient plans in the re-optimized dataset. A closer look at its input and outputs shows that it has the lowest rectal gEUD value of 59.567 Gy with an above average PTV D95 value of 71.725 Gy. This results in the highest output to input ratio of 1.204 within the dataset and explains why the plan is used as a peer for all inefficient plans.

Table 4. Peer counts (the number of times a plan is referred to as a peer) of the efficient plans.

Plan	18	21	25	28	29	31*	32	37
Peer count original	22	6	2	24	30	N/A	2	9
Peer count re-optimized	N/A	23	4	5	27	37	2	N/A

To further justify our approach, we carried out re-optimization for two plans that were characterized as efficient in the initial DEA analysis, plans 28 and 29. The re-optimization was carried out by the same planner who did the re-optimization of the 5 inefficient plans with the same instructions. As mentioned before, the planner was unaware of the results of the DEA analysis.

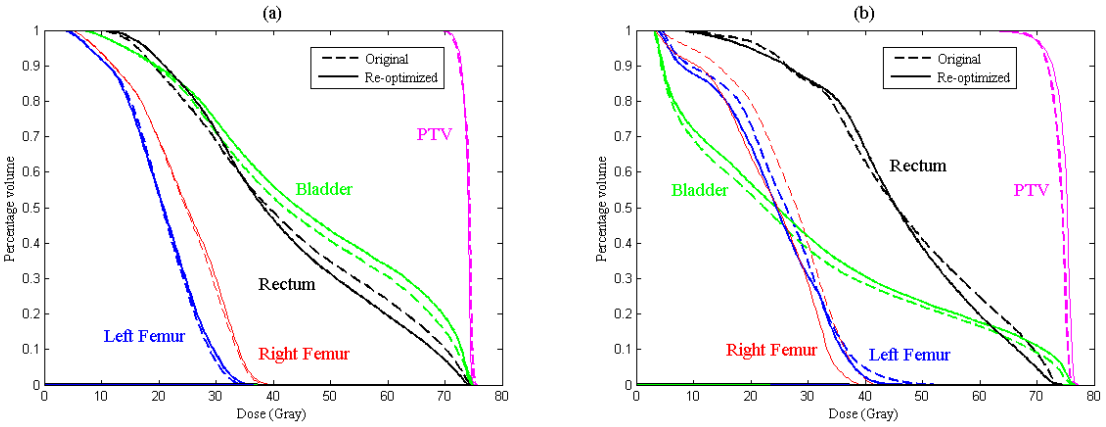

Figure 5. The DVHs of original and re-optimized (a) plans 28 and (b) 29.

The efficiency score of both plans before and after re-optimization is 1. Once again PTV D95 changed by only small amounts (an increase of 0.3 and 0.45 Gy, respectively). Rectal sparing was achieved with a decrease in rectal gEUD of 1.555 and 1.898 Gy, respectively. The DVHs in figure 5 illustrate the plan changes. But in contrast to the re-optimization of inefficient plans, the improvement in rectal sparing was accompanied by an increase of dose to the bladder and other areas beyond the rectum, which worsen the conformity index for plan 28 and significantly increased dose to the small bowel beyond what was considered as clinically acceptable for plan 29. Clinical review of these plans suggests that the re-optimized plans are not better than the original plans. These results further support the correctness of our DEA model in identifying plan improvement potential.

5. Discussion and Conclusions

In this study we investigate the validity of using DEA as a quality assessment tool for radiotherapy planning. We use DEA to assess the quality of 37 prostate plans using an input-oriented VRS model

with rectal gEUD as the input, PTV D95 as the output and the percentage volume of rectum overlapping PTV as a non-discretionary output variable. Five plans that are considered of lesser quality by DEA are re-optimized with the goal of further improving the dose to rectum while maintaining clinical acceptance. After re-optimization, the dose to the rectum for all five plans improved substantially without clinically considerable deterioration in PTV coverage. In addition, the input and outputs of the re-optimized plans are very close to DEA's prediction of the best achievable plan, with a maximum difference of 0.8 Gray. These results confirm that DEA is capable of identifying plan improvement potential and predicting the best attainable plan in terms of the input and outputs.

There are several advantages of using DEA methodology as a quality assessment tool for radiotherapy treatment planning. Firstly, DEA is non-parametric. For non-parametric approaches, there is no need to assume a functional form for the efficient frontier whereas such an assumption is required for parametric approaches. While parametric approaches such as regression analysis can also be used to estimate the efficient frontier, it can be difficult to specify a functional form for the efficient frontier especially when there are multiple inter-related parameters. In contrast, the non-parametric nature of DEA allows practitioners to select the inputs and outputs that are considered most relevant in assessing the quality of radiotherapy treatment plans without too much concern on the underlying relationship among them. The non-parametric nature of DEA leads to its second advantage: the ability to handle multiple inputs and multiple outputs. This ability is particularly important in assessing radiotherapy treatment plans due to the conflicting nature of treatment objectives. Assessing radiotherapy treatment plans based solely on the maximal OAR sparing might encourage the planners to generate plans with maximal OAR sparing but near minimal acceptable PTV coverage. DEA methodology allows plan assessment based on multiple inputs and outputs and therefore captures the conflicting nature of treatment planning more adequately. Thirdly, DEA constructs an efficient frontier based on the best attainable results in the dataset of historical treatment plans. This is distinctly different to ordinary least squares regressions that attempt to fit the regression function at the centre of the data spread and provide estimations for the "average" attainable results rather than the best attainable results. In this radiotherapy application, since we are interested in the best attainable results, we consider DEA to be preferable to regression methods. Fourthly, DEA not only provides the efficiency score for the plan being assessed, but also target information, including the peers and the corresponding weights. The target shows the inputs and outputs required to make a radiotherapy treatment plan efficient. A treatment planner can compare the target with the treatment plan being assessed and decide if further planning is required. If the target is largely composed from a particular peer, a treatment planner can trace back to the peer, assess how the peer is derived and perhaps conduct the re-optimization using similar treatment objectives and/or objective weights. In addition, DEA's capability to accurately predict the best attainable dose for both the PTV and the OAR allows planners to set achievable plan objectives. Wu *et al* (2011) suggested that by setting achievable plan objectives, one can reduce the trial-and-error attempts required to find a satisfactory plan. Last but not least, DEA is readily available in many software packages (Barr 2004) and can be conducted independently of the treatment planning system with negligible computational effort. This provides clinics with an approach to improve planning efficiency and plan quality without making changes to the treatment planning system.

While DEA offers many advantages, it is not without some potential limitations. One limitation is that the efficiency score for a plan is a relative measure compared to other plans in the dataset. Thus a plan rated fully efficient for a dataset might not be truly the best plan, but simply a superior plan compared to the plans in the dataset. However, as more efficient plans are generated and included in the dataset, DEA will be able to learn from the plans and will be able to approximate the true efficient frontier more accurately. As a result, this limitation would become less significant over time. Another limitation is that, as more inputs and outputs are included in the formulation, DEA starts to lose discrimination power on the performance of the DMUs. Introducing more inputs and outputs imposes more constraints in the formulation and thus further constrains the set of potential DMUs. As a consequence, the DMUs are closer to the efficient frontier and more DMUs will be deemed efficient or close to be efficient. To address this limitation, we note that in practice at most two or three clinically relevant trade-offs between the objectives need to be considered (Hoffmann *et al* (2006), see also Holdsworth *et al* (2011)). By including only these relevant objectives in the formulation, the

number of inputs and outputs can be effectively controlled while maintaining the discrimination power. The last limitation is that a plan can be rated efficient simply because it has an optimal value in one of the DEA inputs or outputs compared to all other plans. For example, the plan with the lowest rectal gEUD and the plan with the highest PTV D95 will be rated as efficient regardless the values of other DEA inputs and outputs. This limitation can be alleviated by checking if the plan is referred to as a peer for other plans. In general, given a database of reasonable size, an efficient plan with preferable output to input ratio is likely to be used as a peer for another plan. In contrast, a plan considered efficient simply because it has an optimal value in one of the input/output values is usually not used as a peer for other plans. Thus by checking the peer counts, we can effectively identify efficient plans that may not be truly desirable.

Multi-criteria optimization (MCO) techniques (Thieke *et al* 2007, Shao and Ehrgott 2008, Holdsworth *et al* 2010, Holdsworth *et al* 2012) and prioritized goal programming techniques (Falkinger *et al* 2012, Breedveld *et al* 2009, Wilkens *et al* 2007, Jee *et al* 2007) offer an alternative approach to improve planning efficiency and ensure plan quality. Rather than working towards an ideal defined by historical plans, MCO techniques define an ideal by means of several objective functions. They derive a representative set of plans in one optimization round and allow planners to explore the plans effectively and to select the best available plan among them. Prioritized goal programming techniques conduct a sequence of optimizations in which high priority goals are optimized first and turned into constraints for the subsequent optimization. At the end of the optimization round, a single plan that best matches the treatment priorities is produced. While these techniques can be used to improve planning efficiency and plan quality, they are not yet adopted by the majority of the clinics. Currently MCO is only available in one commercial treatment planning system which is quite new to the market at this stage. Centres who are interested in using it would need to replace their existing system which is quite an effort from both fiscal and implementation viewpoints. In contrast, DEA offers an easily implementable approach to improve the planning efficiency and plan quality, without the need to change the treatment planning system. In addition, it is worth mentioning that the DEA method in radiotherapy plan assessment can be of assistance to these advanced planning techniques. For planning using MCO techniques, the representative set of plans can consist of a large number of plans and it may be difficult for the planner to explore and compare all the plans thoroughly. A well established DEA model will assist the planner to quickly identify plans of best quality from the representative set and thus enable an even more efficient planning process. For planning using prioritized goal programming approaches, since planners are provided with only a single plan that matches the treatment goals as closely as possible, it is important to have additional measures to reassure the quality of the generated plan. In this case, DEA can be used to compare the generated plan to past plans and measure if the plan is truly of high quality.

Further investigation is required to extend the DEA model to include more and/or other plan assessment criteria. In this preliminary study we use PTV D95 and rectal gEUD to account for the dose to the PTV and the rectum, respectively. Planning criteria associated with other OARs such as bladder and femur heads are generally unchallenging and, in our opinion, would not require DEA analysis. It may, however, in the case of prostate planning be useful to include second PTV or rectal parameters such as PTV conformity index or rectal V70Gy (percentage volume of a structure that receives at least 70 Gy of radiation). As previously discussed, adding to the number of plan assessment criteria in the DEA model degrades plan quality discrimination. In future research, we will investigate the most effective plan assessment criteria that should be included in the DEA model, followed by an investigation on how these plan assessment criteria can be incorporated in the DEA model while maintaining sufficient discrimination on the quality of the plans. Other treatment sites likely to require a larger number of inputs and outputs such as head and neck will also be investigated in future work.

Acknowledgement

We thank Dayan Loria from Auckland Radiation Oncology for conducting the re-optimization. The research leading to these results has received funding from the New Zealand Government and European Union Seventh Framework Programme (FP7-PEOPLE-2009-IRSES) under grant number 246647 (project OptALI).

References

Ågren-Cronqvist A-K 1995 *Quantification of the Response of Heterogeneous Tumours and Organized Normal Tissues to Fractionated Radiotherapy* PhD thesis Stockholm University

Banker R D and Morey R C 1986 Efficiency analysis for exogenously fixed inputs and outputs *Operations Research* **34** 513-21

Barr R S 2004 DEA software tools and technology. *Handbook on Data Envelopment Analysis* ed W W Cooper *et al* (New York: Springer US)

Breedveld S, Storchi P R M and Heijmen B J M 2009 The equivalence of multi-criteria methods for radiotherapy plan optimization *Phys. Med. Biol.***54** 7199-209

Charnes A, Cooper W W and Rhodes E 1978 Measuring the efficiency of decision making units *European Journal of Operational Research* **2** 429-44

Cheung R, Tucker S L, Lee A K, De Crevoisier R, Dong L, Kamat A, Pisters L and Kuban D 2005 Dose-response characteristics of low- and intermediate-risk prostate cancer treated with external beam radiotherapy *Int. J. Radiation Oncology Biol. Phys.***61** 993-1002

Chilingerian J A and Sherman H D 2011 Health-care applications: from hospitals to physicians, from productive efficiency to quality frontiers *Handbook on Data Envelopment Analysis* ed W W Cooper *et al* (New York: Springer US)

Coelli T J, Rao D S P, O'Donnell C J and Battese G E 2005 *An Introduction to Efficiency and Productivity Analysis* (New York: Springer US)

Cooper W W, Seiford L M and Zhu J 2011 Data envelopment analysis: history, models and Interpretations *Handbook on Data Envelopment Analysis* (New York: Springer US)

Deasy J O, Blanco A I and Clark V H 2003 CERR: A computational environment for radiotherapy research *Med. Phy.* **30** 979-85

Falkinger M, Schell S, Müller J and Wilkens J J 2012 Prioritized optimization in intensity modulated proton therapy *Z. Med. Phy.* **22** 21-28

Hoffmann A L, Siem A Y D, Den Hertog D, Kaanders J H A M and Huizenga H 2006 Derivative-free generation and interpolation of convex Pareto optimal IMRT plans *Phys. Med. Biol.***51** 6349-69

Holdsworth C, Kim M, Liao J and Phillips M H 2010 A hierarchical evolutionary algorithm for multiobjective optimization in IMRT *Med. Phy.* **37** 4986-97

Holdsworth C, Kim M, Liao J and Phillips M H 2012 The use of a multiobjective evolutionary algorithm to increase flexibility in the search for better IMRT plans *Med. Phys*. **39** 2261-74

Holdsworth C, Stewart R D, Kim M, Liao J and Phillips M H 2011 Investigation of effective decision criteria for multiobjective optimization in IMRT *Med. Phy.* **38** 2964-74

Hunt M A, Jackson A, Narayana A and Lee N 2006 Geometric factors influencing dosimetric sparing of the parotid glands using IMRT *Int. J. Radiation Oncology Biol. Phys.***66** 296-304

Jee K-W, Mcshan D L and Fraass B A 2007 Lexicographic ordering: Intuitive multicriteria optimization for IMRT *Phys. Med. Biol.***52** 1845-61

Kallman P, Lind B K and Brahme A 1992 An algorithm for maximizing the probability of complication-free tumour control in radiation therapy *Phys. Med. Biol.***37** 871-90

Lin K-M, Simpson J, Sasso G, Raith A and Ehrgott M 2012 An application of data envelopment analysis to external radiotherapy treatment planning *Proc. Of the 46th Annual Conf. of the Operations Research Society of New Zealand* ed J Haywood 2012 Wellington pp 185-93

Mackie T R, Holmes T, Swerdloff S, Reckwerdt P, Deasy J O, Yang J, Paliwal B and Kinsella T 1993 Tomotherapy: A new concept for the delivery of dynamic conformal radiotherapy *Med. Phy.* **20** 1709-19

Moore K L, Brame R S, Low D A and Mutic S 2011 Experience-based quality control of clinical intensity-modulated radiotherapy planning *Int. J. Radiation Oncology Biol. Phys.***81** 545-51

Niemierko A 1999 A generalized concept of equivalent uniform dose *Med. Phy.* **26** 1100

Phillips M H and Holdsworth C 2011 When is better best? A multiobjective perspective *Med. Phys.* **38** 1635-40

Raith A, Harton K, Lee A, Priddey H and Rouse M 2012 pyDEA - a software package and user interface for DEA *Department of Engineering Science, The University of Auckland*

Ramer R, Holder A and Papanikolaou N 2008 Utilization of data envelopment analysis (DEA) to compare prostate treatment options *Med.Phys.* **35** 2815

Santos S P and Amado C A F 2012 Using data envelopment analysis for formative evaluation of radiotherapy services: An exploratory study. *Advanced Decision Making Methods Applied to Health Care* ed E Tànfani and A Testi (Verlag Italia: Springer Milan)

Shao L and Ehrgott M 2008 Approximately solving multiobjective linear programmes in objective space and an application in radiotherapy treatment planning *Math. Meth. Oper. Res.* **68** 257-76

Shepard D M and Cao D 2011 Clinical implementation of intensity-modulated arc therapy *IMRT, IGRT, SBRT : Advances in the Treatment Planning and Delivery of Radiotherapy (Front. Radiat. Ther. Oncol.* vol 43*)* ed J L Meyer (Basel: Karger) pp 80-98

Thieke C, Küfer K-H, Monz M, Scherrer A, Alonso F, Oelfke U, Huber P E, Debus J and Bortfeld T 2007 A new concept for interactive radiotherapy planning with multicriteria optimization: first clinical evaluation *Radiotherapy and Oncology* **85** 292-98

Wilkens J J, Alaly J R, Zakarian K, Thorstad W L and Deasy J O 2007 IMRT treatment planning based on prioritizing prescription goals *Phys. Med. Biol.***52** 1675-92

Wu B, Ricchetti F, Sanguineti G, Kazhdan M, Simari P, Chuang M, Taylor R, Jacques R and McNutt T 2009 Patient geometry-driven information retrieval for IMRT treatment plan quality control *Med. Phy.* **36** 5497-505

Wu B, Ricchetti F, Sanguineti G, Kazhdan M, Simari P, Jacques R, Taylor R and McNutt T 2011 Data-driven approach to generating achievable dose-volume histogram objectives in intensity-modulated radiotherapy planning *Int. J. Radiation Oncology Biol. Phys.***79** 1241-247

Xing L, Thorndyke B, Schreibmann E, Yang Y, Li T-F, Kim G-Y, Luxton G and Koong A 2006 Overview of image-guided radiation therapy *Medical Dosimetry* **31** 91-112

Yu C X and Tang G 2011 Intensity-modulated arc therapy: principles, technologies and clinical implementation *Phys. Med. Biol.***56** R31-R54

Zhu X, Ge Y, Li T, Thongphiew D, Yin F-F and Wu Q J 2011a A planning quality evaluation tool for prostate adaptive IMRT based on machine learning *Med. Phy.* **38** 719-26

Zhu X, Li T, Yin F-F, Wu Q J and Ge Y 2011b Response to comment on: A planning quality evaluation tool for prostate adaptive IMRT based on machine learning' Med. Phys. 38, 719 (2011) *Med. Phy.* **38** 2821