

Considerations for using data envelopment analysis for the assessment of radiotherapy treatment plan quality

Abstract

Data envelopment analysis (DEA) is a widely used method in operations research for the benchmarking and empirical assessment of productive efficiency. We have previously applied DEA for treatment plan analysis and demonstrated its ability to determine relative plan quality, however considerations regarding the optimal use of DEA were not considered in that work. In the current work we have extended the complexity of the DEA modelling to include an increased number of measures of treatment plan quality as well investigating the best method of accounting for patient geometry. Forty-one IMRT prostate treatment plans were retrospectively analysed using an input-oriented variable returns to scale DEA method. The impacts of DEA weight restrictions were analysed with reference to the ability of DEA to differentiate plan performance at a level of clinical significance. Patient geometry significantly influences plan quality and alternative methods for considering geometry in the DEA model were investigated. In this work we identify how best to use DEA for the relative assessment of prostate treatment plan quality.

1. Introduction

Data Envelopment Analysis (DEA) is a method for benchmarking organisations that constructs a ‘best-practice frontier’ which is used as a comparator for non-best practice organisations. In this paper, radiotherapy treatment plans are used instead of organisations to determine relative plan quality as described in our previous work. In the following we first summarise this work and provide an introduction to DEA in the context of treatment planning. We then analyse forty-one treatment plans retrospectively to identify an appropriate DEA model that can be applied in a clinical context to ensure treatment plan quality.

Previously we used a simple single input and single output model to show that the operations research method of data envelopment analysis (DEA) could be used to assess relative quality among prostate treatment plans (Lin *et al* 2013). In that work we demonstrated that DEA efficiency could be used to assess the relative quality of prostate radiotherapy treatment plans using a single input and output and considering patient geometry via rectum-planning target overlap. We further demonstrated

that replanning low DEA efficiency plans with the aim of increasing the DEA efficiency improved the clinical assessment of plan quality. This paper extends our previous work to multiple inputs whilst introducing more sophisticated models to address different levels of complexity among inputs and investigating different approaches to handling patient geometry. The improved understanding of the interrelationships between DEA inputs, outputs, patient geometry and plan quality is a prerequisite for expanding the role of DEA to other clinical sites and is the aim of this work.

DEA is a widely used method in operations research and management science for assessing the performance of a set of decision-making units (DMUs). In DEA terminology, the DMU is an entity that is responsible for converting inputs into outputs and whose performance is being evaluated (Charnes *et al* 1978, Cooper *et al* 2011). In a loose economic interpretation, the inputs represent the cost we pay to produce outputs. The concept of DEA is directly applicable to the problem of assessing treatment plan quality in radiotherapy in which the doses to organs at risk (OARs) are considered as the cost we pay for delivering dose to the planning target volume (PTV). A DMU in this context is therefore a treatment plan.

DEA is attractive for plan analysis for three reasons: Firstly DEA is data oriented where performance is based only on the observed data with minimal assumptions; secondly DEA performs peer evaluation by comparing treatment plans with respect to a best practice reference set defined by historical plans. In other words it is a method directed towards Pareto efficient frontier generation in contrast to some knowledge based treatment planning solutions that are based on statistical measures; and thirdly the data used for inputs and outputs do not have to be commensurable, i.e. do not have to have a common measure. Thus we can simultaneously use, for example, volume, dose, biological dose, ratios or any other relevant parameter as our inputs and outputs.

DMUs are assessed based on their ability to convert a set of inputs into outputs. This ability is referred to as efficiency and measured as a ratio of weighted outputs and weighted inputs. Given a set of DMUs (in our case, treatment plans), DEA assesses the relative efficiency of each plan by comparing it to all other plans in the set and searching for evidence of potential improvements, i.e. whether some of the plan parameters, cast as DEA inputs and outputs, can be improved without worsening any of its other inputs or outputs. DEA maximises the efficiency rating of each plan by automatically adjusting the weighting of the inputs and outputs to make the plan under investigation appear most favourable. These weights are referred to throughout the paper as DEA input/output weights. If, even under this most favourable consideration, other plans have better efficiency, then the plan under investigation is inefficient. Otherwise, if there is no evidence for potential improvements, the plan is considered fully efficient. If potential improvements are identified then a target solution is determined based on those fully efficient plans (known as peers) that are most similar to the plan being evaluated. The method ensures that the resulting relative efficiency lies between 0 and 1, where 1 is fully efficient. In the following we present three simple examples to illustrate the intuition underlying an application of DEA in the assessment of treatment plan quality.

Example 1: Consider four treatment plans (plan A, B, C, D), each being a historical 78Gy IMRT prostate treatment plan that was delivered to treat a patient in the past. We also consider a fifth plan, plan N, which corresponds to a plan for a new patient that is under evaluation. In the simplest case, similar to what is described by Lin (2013), we consider only two metrics to assess plan quality: The metrics selected are PTV V95, the percentage of PTV volume receiving at least 95% of the prescribed dose and rectal generalised equivalent uniform dose (gEUD). Since the aim is to obtain plans with consistently high PTV V95 and low rectal gEUD, an input-oriented DEA model is chosen which seeks to minimise input (rectal gEUD) while maintaining output (PTV V95). Alternatively, an output-oriented DEA model would seek to maximise output while maintaining input levels. Input-orientation is used in the examples, and justified later. Figure 1 shows rectal gEUD (horizontal axis) and PTV V95 (vertical axis) for the five plans under consideration. Plans A, C and D are DEA-efficient as, within the set of plans considered, there is no evidence that lower rectal gEUD can be achieved while maintaining the same level of PTV V95. The three DEA efficient plans define the Pareto efficient frontier shown as the solid piecewise linear line through plans D, A and C. Those plans have an output to input ratio superior to historical plan B which is not DEA-efficient. The new plan N is also not DEA efficient and the DEA evaluation determines that it should be able to achieve a rectal gEUD

reduction of almost 3 Gy (53.95-56.91) as shown in point N* while keeping the same PTV V95 value (98.39). N* is called the target for plan N and the difference between the two is referred to as Radial Slack. Radial slack is utilised in this work in one of the methods for managing patient geometry and is described later in this paper. DEA also determines the peer plans for N to be A and D as they define the line segment on which target N* lies, and provides information on the relative influence of each peer plan in setting the target.

In the example plans, A and D have a peer-weight of 0.34 and 0.66 respectively and the DEA results are shown in Table 1. The targets for plan N are determined by applying these peer-weights so that rectal gEUD equals 53.95 ($0.34 \times 56.97 + 0.66 \times 52.39$) and PTV V95 equals its actual measure 98.39 ($0.34 \times 99.05 + 0.66 \times 98.05$). Peer plans indicate what a plan (such as N) should be able to achieve and may also help a planner determine how this could be achieved by considering comparative, but better, plans determined by DEA. The DEA model employed here is a so-called variable returns to scale model which appropriately models trading off changes in inputs and outputs in radiotherapy planning. The alternative would be a constant returns to scale model which assumes there is a constant relationship between an increase of input and output. In Example 1 this would mean that a unit increase of rectal gEUD would always correspond to a constant increase of PTV V95. However, the relationship between rectal gEUD and PTV V95 varies in a complex manner and hence a variable returns to scale DEA model is used in the following examples and results. Example 2 generalises the DEA concept to a case with two inputs.

Example 2: In addition to the data used in Example 1 we now consider a second input, PTV dose max, the maximum dose to PTV, as discussed later. Example 2 also includes one additional plan, plan E. In Example 2 we also assume that PTV V95 values are identical for all plans. This assumption is not a DEA requirement, but is made to be able to illustrate the problem in the two input dimensions only. Figure 2 illustrates the two inputs for all six considered plans. DEA now evaluates plan efficiency as the ratio of weighted sum of outputs and weighted sum of inputs. The weights determined are those most favourable for the plan under consideration, taking into consideration all other plans. In radiotherapy planning this means that DEA makes no *a priori* assumption of the relative importance of one input or output over another. In other words a DEA model recognises that different trade-offs are made for different patients. Table 2 shows an example where three different sets of weights are applied to the plan data. For weights $w_1 = 0.0$ and $w_2 = 1.0$, plan A is found to be DEA efficient; for weights $w_1 = 0.08$ and $w_2 = 0.92$ plans A and E are DEA efficient. For weights $w_1 = 0.24$ and $w_2 = 0.76$, both plans D and E are DEA efficient. In fact, plans A, D and E are the only efficient plans in the example. For the others, plan N for example, its efficiency score is 0.979 ($w_1 = 0.0$ and $w_2 = 1.00$), 0.980 ($w_1 = 0.08$ and $w_2 = 0.92$), 0.976 ($w_1 = 0.24$ and $w_2 = 0.76$), and 0.920 ($w_1 = 1.0$ and $w_2 = 0.0$). Thus, the best weights for plan N are $w_1 = 0.08$ and $w_2 = 0.92$, and it obtains an efficiency score of 0.980 which is its best efficiency score calculated for its best weights.

Without weight restrictions, the peer plans for plan N are A and E, where plan A is a slightly more important peer with a peer-weight of 0.54. DEA determines that the relative importance (or weight) put on rectal gEUD versus PTV dose max is 0.08 and 0.92 for plan N, respectively. Most other plans, with the exception of plan D, have similar optimal weights. Only plan D has the option to put all the weight on rectal gEUD. This result would indicate that four of the considered plans look most favourable when PTV dose max is considered and not when considering rectal gEUD, given that DEA is allowed to choose weights without restriction. This may be an undesired result as rectal gEUD is arguably the more important factor here. DEA allows us to add weight restrictions to control this, e.g. we could require that the DEA weight on rectal gEUD must be at least that of the weight on PTV dose max, leading to a different, but more clinically desirable assessment of plan efficiency scores and targets. With the weight restriction, plan N now has optimal input weights of $w_1 = 0.50$ (rectal gEUD) and $w_2 = 0.50$ (PTV dose max), and a new target point N** in Figure 2. The choice of weight restrictions affects the target. If, for instance the weight of rectal gEUD is only required to be at least a fifth of the weight of PTV dose max, the target for N coincides with point E in Figure 2 (this is not shown in the figure). The introduction of weight restrictions also renders plans inefficient that were initially efficient, such as plan A which becomes inefficient in both cases.

Example 3: We now assume the same setting as in Example 1 with only two parameters to assess plan quality (rectal gEUD and PTV V95). Comparing plans as outlined in Example 1 is not entirely fair, as Example 1 disregards differences in patient geometry. When considering rectal gEUD, one factor that strongly affects this value is the patient-specific overlap between PTV and rectum, which is out of the radiotherapy planner's control. In Figure 3 this overlap is indicated by the size of the circle where bigger means higher overlap. When assessing quality of plan N via DEA efficiency, it does not appear fair to compare plan N which has a relatively high overlap of 12.43% to A and D which both have lower overlap. Hence, the target determined for N in Example 1 may not be achievable.

When performing DEA, one may want to take environmental factors into account. Environmental factors are factors that influence the performance of DMUs but are external to and out of the control of the DMUs. Patient anatomy or geometry is a treatment planning environmental factor since the relative positions of PTV and organs at risk clearly affect plan quality yet it cannot be influenced by the planning process. In the case of prostate radiotherapy planning, the percentage overlap of PTV and rectum is the dominant environmental factor. How to best account for patient geometry within the DEA process is discussed later.

In summary, Example 1 is straightforward as it has a single input and output. Efficiency could have been determined as a simple normalised ratio of output over input. Likewise, Example 2 was used to illustrate the use of input and output weights in DEA. Finally, Example 3 introduces the idea of environmental factors outside of a planner's control. We further expand on all these ideas in the following. It is important to note that DEA can cope with any number of inputs and outputs in principle but its discriminatory ability can decline with increasing numbers of inputs and outputs especially with small data sets. DEA automatically determines the weights by solving an optimisation problem for each plan under investigation, thereby applying appropriate operations research techniques. It may be useful at this point to introduce the mathematical description of DEA using variable returns to scale. For more detailed descriptions the reader is referred to Charnes *et al* (1994). The DEA efficiency of plan k from $j=1, \dots, n$ plans is given by solving the following linear programme (model stated in multiplier form and variable returns to scale without consideration of weight restrictions or environmental variables):

$$\text{Max } h_k = \sum_{r=1}^s u_r y_{rk} - u_0 \quad (1)$$

Subject to

$$\sum_{i=1}^m v_i x_{ik} = 1$$

$$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} - u_0 \leq 0, \quad (j = 1, \dots, n)$$

$$u_r \geq 0, \quad (r = 1, \dots, s)$$

$$v_i \geq 0, \quad (i = 1, \dots, m)$$

Where:

h_j is the relative efficiency of plan j

v_i is the weight determined for input i ;

m is the number of inputs;

u_r is the weight variable determined for output r ;

u_0 is a variable related to returns to scale;

s is the number of outputs;

x_{ij} is the magnitude of the i -type input for plan j ;

y_{rj} is the magnitude of the r -type output for plan j ;

n is the number of plans.

Additional constraints can be added to the basic DEA model to ensure adequate consideration of environmental variables and weight restrictions. Several forms of weight restrictions can be included (see also Allen *et al* 1997). Here we apply a constraint to ensure that the ratio of weights of two inputs, i_1 and i_2 , is at least α : $\frac{v_{i_1}}{v_{i_2}} \geq \alpha$ or equivalently $v_{i_1} \geq \alpha v_{i_2}$. In Example 2 we would choose $\alpha = 1$. To capture the environmental factor in Example 3, it can be included as another output to the model since this ensures the comparison target is computed with at least the same overlap or the model can be modified to take environmental differences into account as explained below.

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. As with our previous work we focus on input-oriented models which make sense in the context of treatment plan design since it is usual to optimise the plan to ensure that the PTV receives the prescribed dose after which potential reductions of dose to organs at risk are explored. An input oriented model provides the target doses, that is the minimum doses to OARs, which should be achievable by the treatment plan.

2. Method

This retrospective non-interventional study was subject to Institutional ethics approval. Forty one, 78Gy IMRT prostate treatment plans were selected based only on a date range. All plans were generated using an Eclipse treatment planning system V11 (Varian Medical Systems, Ca, USA) and all were clinically acceptable. The twin treatment planning requirements of PTV cover and minimisation of normal tissue dose was considered in the DEA plan evaluation. The metrics selected were PTV V95, the percentage of PTV volume receiving at least 95% of the prescribed dose and rectal generalised equivalent uniform dose (gEUD) with the volume effect parameter equal to 8 (Niemierko 1997). In practical terms, the rectum is the only normal tissue structure that impacts prostate PTV cover and therefore the only normal tissue structure considered in this analysis. Other organs at risk can easily be accommodated by the DEA method, which is a strength of DEA. However, a greater number of inputs may result in a greater number of fully efficient plans. It is therefore sensible to limit the organs at risk to only those that most influence plan quality, especially when only a low number of plans is analysed. In treatment planning it is usual to constrain the maximum dose to the PTV; therefore a second input of PTV maximum is used. The key summary statistics for the plans used are; rectal gEUD mean 59.2 Gy (range 52.4-62.1 Gy), PTV-rectal overlap 11.7% (5.5-18.1 %), PTV V95 98.2 % (95.1-99.9 %) and PTV dose max 82.0 Gy (80.2-84.5 Gy), shown in Figure 4. As with our previous work, an input orientated variable returns to scale approach was used. In order to determine the most suitable DEA model, it is necessary to understand how the handling of two parameters, patient geometry and DEA weights, can influence the outcome.

The planning trade-off between dose to PTV and rectum is influenced by the patient geometry and in particular the overlapping volume of rectum and PTV. This may be considered an environmental factor in DEA terms and can be handled in one of three ways, as either a DEA output (as this ensures the DEA model targets have the same or worse environmental factor), by categorisation or by making an overlap-based adjustment to the rectal dose. The first approach would thus have two outputs, PTV V95 and the overlap, and two inputs gEUD (Gy) and PTV dose max (Gy) (models M1 and M4 in Table 3). Since we have assumed an input orientation and variable returns to scale, these models identify the minimum level of input associated with the respective levels of output. Thus the DEA model would ensure that plans with high (or low) levels of overlap output are only compared with target plans with equal or greater levels of overlap. The second approach, categorisation (Banker and Morey, 1986), uses the amount of overlap to construct an ordinal ranking

of plans from low to high overlap so that high overlap plans are only compared with similar plans whereas low overlap plans are compared with all other plans including higher levels of overlap (M2 and M5 in Table 3). The third approach adjusts the rectal dose taking overlap into consideration using a three stage process comprising firstly a DEA analysis without considering overlap, secondly construction of a different frontier of radial slack from the first stage to identify the minimum amount of slack due to overlap, and thirdly rerunning a DEA analysis but with the inputs adjusted for this minimum amount of slack due to overlap (M3 and M6).

To evaluate the influence of DEA weights, the three approaches just cited were repeated with the addition of relative DEA input weight restrictions to ensure that the rectal dose is appropriately prioritised over the PTV dose max. In the repeat models we constrained the ratio of rectal dose to PTV dose max DEA input weighting to be ≥ 10 . This value was chosen based on its ability to identify potential rectal dose improvements at a level of clinical significance as discussed later. Table 3 shows the attributes of the six models.

Model names indicate model setup, as explained in this paragraph: E stands for environmental, with options output (O), categorical (C) and adjusted (A), and W for weight restrictions with options yes (Y) and no (N). In models M1: EO_WN and M4: EO_WY the percentage of PTV overlapping the rectum is considered as a second DEA output. In models M2: EC_WN and M5: EC_WY the overlap is considered by placing a plan in one of three categories according to degree of overlap. A plan is then only compared to plans with the same or higher overlap category. In models M3: EA_WN and M6: EA_WY the DEA target values are adjusted according to the observed limiting influence of overlap as explained in the next paragraph. In models M1: EO_WN, T2: EC_WN and M3: EA_WN, DEA was performed without constraining the input weights, whilst in models M4: EO_WY, M5: EC_WY and M6: EA_WY, the relative input weights were constrained.

In DEA, radial slack is a measure of the inefficiency of production, in an economical interpretation, and in our application is the extent to which it might be possible to reduce doses to the organ at risk (rectum). Radial slack can be considered to be comprised of two components, that due to patient geometry (PTV-rectal overlap) and that due to plan quality. To evaluate radial slack we first perform DEA on the dataset without considering overlap. Second, the resultant radial slack for each plan is plotted against its overlap from which a frontier is determined by a linear fit of points on the Pareto efficient frontier. The frontier effectively describes the minimum amount of radial slack attributable to overlap. With reference to Figure 5, two plans with a similar PTV-rectal overlap of approximately 12.7% are highlighted. One lies on the overlap frontier with a radial slack value of 2.3 Gy, wholly attributable to overlap. The other has a total radial slack of 8.1 Gy of which $8.1 - 2.3 = 5.8$ Gy is therefore attributable to plan quality. Since the overlap component of radial slack cannot be influenced by the planning process, for each plan this is subtracted from the rectal gEUD to yield the planning component of radial slack. We can refer to this as the adjusted rectal gEUD. To calculate DEA efficiency, rectal gEUD is replaced with the adjusted rectal gEUD without the requirement to further consider overlap.

The Spearman rank order correlation was used to compare the results of each model to assess if the ranking order was affected by the trial method (SigmaPlot 13.0, Systat Software Inc. CA). A comparison of plans appearing within the upper and lower quartiles of efficiency score for each model was also undertaken to evaluate inter-model consistency. An in-house software package pyDEA (Raith et al.2012) was used to perform the DEA analysis.

3. Results

The calculated efficiency score and range of efficiency scores is dependent upon the DEA model used and is shown in Figure 6. The greatest range was observed when the environmental factor is incorporated into the adjusted rectal dose (M6:EA_WY). This is in part due to the much reduced scale

of the rectal dose input data. This is not seen in its non-weight constrained version (M3:EA_WN) because the dominant input in this model is the maximum dose to PTV (PTV dose max). The other two weight constrained models (M4:EO_WY and M5:EC_WY) displayed a greater range of efficiency scores than any of the unconstrained models due to the increased significance of the rectal dose over PTV dose maximum. Of greater importance than the range of efficiency scores, however, is the ability of the DEA model to identify the plans that constitute the efficient frontier and those which can significantly improve. Thus the ranking order of plans is a more important consideration than the efficiency range. The Spearman rank order correlation indicates that there is a positive correlation between the relative rank of individual plans obtained by any of the models compared to any other and that the relationship is statistically significant ($P < 0.05$) (Table 3). There is a strong association between the three unrestricted weight models M1:EO_WN, M2:EC_WN and M3:EC_WN and also between the three restricted weight models M4:EO_WY, M5:EC_WY and M6:EA_WY but a weak association between restricted and unrestricted weight models.

The number of fully efficient plans ranged from four to nine dependent on the model and three plans were rated fully efficient by all six models. There was a difference in the identification of the most inefficient plans between the three models not using weight restrictions and the three models using weighting restrictions. In both cases, seven plans were determined to lie within the lowest efficiency quartile irrespective of which of the three variants was used. However of those seven plans, only two were common to both the groups without weight restriction and with weight restriction. Four plans were ranked in either the upper and lower quartiles dependant on the DEA model. Two of these plans were assessed as maximally efficient when no weight restrictions were used though possible rectal improvements were identified with the addition of weight restrictions.

Target doses for rectum and PTV dose max were influenced by the addition of the relative DEA weight restriction. Without DEA weight restrictions, the mean (standard deviation) of identified possible reduction to rectal gEUD and PTV maximum doses was 0.86 (0.73) and 1.12 (0.90) Gy respectively. With the introduction of the weight restriction the identified possible reduction to rectal gEUD and PTV maximum doses was 3.13 (2.20) and 0.98 (0.97) Gy respectively. Thus the addition of weight restrictions allowed for the identification of significant potential rectal improvements at the expense of a small reduction in PTV maximum identification as shown in Figure 7.

4. Discussion

The strength of DEA lies in a minimal number of assumptions regarding the relationships between the inputs and outputs and estimation of DEA weights, which can indicate the relative importance of different parameters when it comes to DEA performance. It is important, however, that the DEA method chosen reflects the most important treatment planning priorities. In other words, it would not be helpful if the efficiency ranking was sensitive to clinically insignificant differences in plan quality. For example, it would be undesirable for a plan to be identified as DEA efficient due to having a small PTV dose maximum when the rectal dose could be significantly further reduced, as was the case for some plans here when using unrestricted DEA weights. With our data it was found that restricting the allowable ratio of rectal gEUD to PTV maximum weighting we could limit the influence of relatively small differences in PTV maximum whilst providing a suitable emphasis on clinically significant rectal dose reduction. To validate this choice of weight restriction differing weight limits were applied to one of the models (model M4: EO_WY that treated overlap as an output) to assess the effect on rectal and PTV maximum dose targets. Weight limits used for the ratio of weights were: none, ≥ 0.2 , ≥ 0.5 , ≥ 1 and ≥ 10 . It was found that the choice of weight limit did not affect the target doses provided that it was ≥ 0.5 as shown in Figure 8. The probability of rectal toxicity increases by the order of 1% per Gy increase in rectal gEUD at this dose level (Tucker *et al.*, 2012) and hence to be clinically useful the DEA method needs to be sensitive enough to identify those plans where

improvements to rectal gEUD greater than several Gy are possible. Figure 7 shows box plots of identified potential dose reductions to rectum and PTV maximum for each of the six models. This demonstrates that unless appropriate weight restrictions are used as in models M4:EO_WY, M5:EC_WY and M6:EA_WY, the DEA model with two inputs is unable to identify plans for which significant rectal dose reduction is possible. It also demonstrates that the cost of using weight restrictions is small in terms of potential reduction to the PTV maximum doses though it is observed that there is a target dose reduction for all models with smaller dispersion of doses. It is clear therefore that a weight restriction is required to ensure clinical relevance when multiple DEA inputs are used and further research is required to determine the best choice of weight restrictions for differing organs at risk under consideration.

How to handle the environmental factor (geometric descriptor) is considered by way of three possible approaches; by considering rectal-PTV overlap as an output, by categorisation (i.e. placing the plans into a number of bins dependent on their overlap) or by adjusting the rectal target dose based on overlap. All three models have a strong statistical association. However, categorisation suffers from the somewhat arbitrary choice of category boundary and the fact that at each category boundary a plan could be fully efficient simply because it is at the boundary. Hence it is possible that a plan that could be significantly improved would not be identified using a categorical approach. Of the other two models, the plan ranking is shown to be very similar as shown in Figure 9, though treating the overlap as an output has the advantage of requiring less data manipulation. Notwithstanding, the adjusted dose method identifies the amount of radial slack (reduction in gEUD) attributable to the amount of overlap which potentially could stimulate better understanding of the relationship between these two variables in further efficiency studies.

5. Conclusion

The input oriented DEA method is the method of choice when performing plan analysis using the key undesirable plan metrics as the DEA inputs. When considering multiple inputs, it is necessary to constrain the DEA input weights in order to identify potential plan improvements at a level of clinical significance. In this case we wished to identify rectal dose improvements of greater than 2 Gy which was only possible using a relative rectal to PTV maximum DEA input weight restriction of at least greater than 0.5. There appears to be no clear advantage for any given approach for the handling of the geometric consideration PTV-rectal overlap; though the categorisation method is somewhat arbitrary.

The flexibility of DEA in modelling the appraisal of plans has been demonstrated in this paper as has been the ability to incorporate clinical knowledge and expectations into the analysis. The use of DEA to identify plans that are best practice to form an ongoing benchmark reference set provides a new framework for clinicians and researchers alike. Over time, one might expect this reference set to change as new technologies and thinking emerge and an obvious question is whether the frontier is moving outwards and at what rate.

Due to the sensitive nature of the data, we are not in a position to share data as we do not have permission to do so.

References

- Allen R, Athanassopoulos A, Dyson RG, and Thanassoulis E (1997) Weights restrictions and value judgements in Data Envelopment Analysis: Evolutions, development and future directions *Ann. Oper. Res.* **73** 13-34
- Banker RD and Morey RC (1986) The use of categorical variables in data envelopment analysis *Manag Sci* **32** 1613-1627
- Charnes A, Cooper W W and Rhodes E (1978) Measuring the efficiency of decision making units *Eur. J. Oper.Res.* **2** 429-44
- Charnes A, Cooper WW, Lewin AY and Seiford LM (1994) Data envelopment analysis: Theory, Methodology, and applications. New York: Springer
- Cooper WW, Seiford L and Mand Zhu J (2011) Data envelopment analysis: history, models and interpretations *Handbook on Data Envelopment Analysis* (New York: Springer)
- Lin K-M, Simpson J, Sasso G, Raith A and Ehrigott M (2013) Quality assessment for VMAT prostate radiotherapy planning based on data envelopment analysis *Phys. Med. Biol.* **58** 5753-5769
- Niemierko A (1997) Reporting and analyzing dose distributions: A concept of equivalent uniform dose *Med Phys* **24** 103-110
- Raith A Harton K Lee A Priddey H and Rouse M (2012) pyDEA—a software package and user interface for DEA. Auckland: Department of Engineering Science, The University of Auckland
- Tucker SL, Dong L, Michalski JM, Bosch WR, Winter K, Cox JD, Purdy JA, and Mohan R (2012) Do intermediate radiation doses contribute to late rectal toxicity? An analysis of data from RTOG 94-06 *Int J Radiat Oncol Biol Phys.* **84** 390-395