Physics and Imaging in Radiation Oncology 3 (2017) 53–56

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.elsevier.com/locate/phro

Original Research Article

A clinical distance measure for evaluating treatment plan quality difference with Pareto fronts in radiotherapy

Kristoffer Petersson[a], Archonteia Kyroudi[a], Jean Bourhis[b], Crister Ceberg[c], Tommy Knöös[c,d], François Bochud[a], Raphaël Moeckli[a,*]

[a] Institute of Radiation Physics (IRA), Lausanne University Hospital, Lausanne, Switzerland
[b] Department of Radiation Oncology, Lausanne University Hospital, Lausanne, Switzerland
[c] Department of Medical Radiation Physics, Clinical Sciences, Lund, Lund University, Lund, Sweden
[d] Radiation Physics, Skåne University Hospital, Lund University, Lund, Sweden

ARTICLE INFO

Keywords:
Pareto
Plan quality
Clinical grading analysis
Prostate

ABSTRACT

We present a clinical distance measure for Pareto front evaluation studies in radiotherapy, which we show strongly correlates (r = 0.74 and 0.90) with clinical plan quality evaluation. For five prostate cases, sub-optimal treatment plans located at a clinical distance value of > 0.32 (0.28–0.35) from fronts of Pareto optimal plans, were assessed to be of lower plan quality by our (12) observers ($p < .05$). In conclusion, the clinical distance measure can be used to determine if the difference between a front and a given plan (or between different fronts) corresponds to a clinically significant plan quality difference.

1. Introduction

There are several ways of comparing and evaluating treatment plans in radiotherapy [1]. One useful way is based on the Pareto optimality concept [2]. Multi-criteria optimization problems (e.g. treatment plan optimization through inverse planning in radiotherapy) often have a set of solutions, each being an optimal compromise between multiple criteria. These are said to be Pareto optimal as they cannot be dominated, i.e. there is no solution that is better in one criterion without being worse in another [2]. A set of Pareto optimal treatment plans form a Pareto front/surface in criterion space. Pareto front evaluation studies can be performed for multiple evaluation criteria, i.e. in *n*-dimensions (*n*D) [3–5], but are mostly used for treatment technique comparison in two dimensions (2D), exploring the trade-off between two important evaluation parameters (e.g. target coverage vs. mean dose to an organ at risk (OAR)) [6–11]. Previously performed Pareto front evaluation studies have shown the usefulness of this method for comparing and demonstrating dosimetric differences between treatment techniques and/or plan optimization techniques. However, these studies are more qualitative than quantitative, i.e. it has not been possible to conclude if differences found were of clinical and/or statistical significance [6–11]. These studies were also limited to evaluating an *n*D problem in 2D, as it is challenging to visualize and evaluate *n*D fronts. These limitations have prevented the method from becoming more widely used within radiotherapy research. Hence, the purpose of our study was to present a measure that mitigates these limitations of previously performed Pareto front studies, i.e. which allows demonstrated differences in 2D/*n*D Pareto front studies to be quantifiable with a clinically relevant measure.

2. Materials and methods

2.1. Introducing the clinical distance measure

The distance between the evaluation parameters of two treatment plans can be calculated with the Pythagorean theorem (see Section A1 in the Supplementary appendix). To add clinical meaning to this purely mathematical distance, we introduced a clinical scaling factor $k = (k_1,k_2,...,k_n)$, which scales the different evaluation parameters. The scaling value was based on the clinical importance of the evaluation parameter and on how difficult it was to influence, e.g. for an OAR evaluation parameter it depended on its volume and proximity to the target(s), increasing with lager volume and proximity to the target(s). The actual scaling values used (Table 1) were extracted from several previously performed Pareto front evaluation studies, of treatments in the brain, head & neck, thoracic, abdominal, and pelvic region [7,11], by quantifying the relative dynamic range between evaluation parameters in the trade-off region (the value range of the trade-off region in one evaluation parameter in relation to the value range in the other evaluation parameter, i.e. how difficult one evaluation parameter was

* Corresponding author at: Institute of Radiation Physics, Rue du Grand-Pré 1, CH-1007 Lausanne, Switzerland.
E-mail address: raphael.moeckli@chuv.ch (R. Moeckli).

http://dx.doi.org/10.1016/j.phro.2017.09.003
Received 24 July 2017; Received in revised form 19 September 2017; Accepted 26 September 2017

Table 1
Scaling values (k_n) coupled with clinical evaluation parameters (n).

k_n	n
1.0 (Gy^{-1})	Maximum dose (Gy) for critical OAR (e.g. spinal cord), CTV coverage
0.5	Min/max constraints (relative volume, %) for PTV (Hot/cold spots, coverage)
0.2 (Gy^{-1})	Constraints for large OAR (Gy) (e.g. mean dose to liver, brain, lungs)
0.1 (Gy^{-1})	Constraints for radiosensitive OAR (Gy) (e.g. mean dose to kidneys, lenses)
0.1	Constraints for OAR in high dose regions (relative volume, %) (e.g. Rectum $V_{90\%}$)
0.05 (Gy^{-1})	Constraints for other OAR (Gy) (e.g. mean dose to rectum, parotid glands)

to influence in relation to the other). Each parameter's clinical importance was also considered so that the values also reflects the clinical priority/preference regarding competing plan objectives [12].

The clinical distance (*cd*) between evaluation parameter values ($1 \leqslant i \leqslant n$) of a dominated treatment plan a and the Pareto front B (consisting of several (M) Pareto optimal treatment plans b) was defined as:

$$\begin{aligned} cd_B(a) &= min_{b\in B}||(a-b)k|| \\ &= min\left(\sqrt{\sum_{i=1}^{n}((a_i-b_{i,1})k_i)^2},\sqrt{\sum_{i=1}^{n}((a_i-b_{i,2})k_i)^2},\ldots \right. \\ &\quad \left. ,\sqrt{\sum_{i=1}^{n}((a_i-b_{i,M})k_i)^2}\right) \end{aligned} \tag{1}$$

Similarly, the clinical distance between two fronts A and B ($a_{i,1},a_{i,2},\ldots,a_{i,L} \in A$ and $b_{i,1},b_{i,2},\ldots,b_{i,M} \in B$) consisting of plans created for two different treatment techniques can be calculated as:

$$\begin{aligned} cd_B(A) &= \frac{1}{L}\cdot\sum_{j=1}^{L}(\pm)cd_B(a_j) = \frac{1}{L}\cdot\sum_{j=1}^{L}(\pm)min_{b\in B}||(a_j-b)k|| \\ &\quad + if\ a_j > b, -if\ a_j < b \end{aligned} \tag{2}$$

If the fronts cross over each other, the difference in plan quality was found by adding the distances for the plans a_j above B and subtracting the distances for the ones below.

A slightly modified version of the clinical distance measure, the plan quality difference measure (*qd*) was also defined. It has the ability to quantify and, if needed, take into account the effect of variations in all dimensions have on the plan quality. The *qd* is calculated as:

$$\begin{aligned} qd_B(a) &= min\left(\sqrt{\sum_{i=1}^{n}((a_i-min(a_i,b_{i,1}))k_i)^2}\right. \\ &\quad -\sqrt{\sum_{i=1}^{n}((b_{i,1}-min(a_i,b_{i,1}))k_i)^2},\sqrt{\sum_{i=1}^{n}((a_i-min(a_i,b_{i,2}))k_i)^2} \\ &\quad -\sqrt{\sum_{i=1}^{n}((b_{i,2}-min(a_i,b_{i,2}))k_i)^2},\ldots,\sqrt{\sum_{i=1}^{n}((a_i-min(a_i,b_{i,M}))k_i)^2} \\ &\quad \left.-\sqrt{\sum_{i=1}^{n}((b_{i,M}-min(a_i,b_{i,M}))k_i)^2}\right) \end{aligned} \tag{3}$$

Eq. (3) is equal to Eq. (1) if all $a > b$, otherwise Eq. (3) $<$ Eq. (1), i.e. $qd \leq cd$.

2.2. Demonstrating the functionality of the clinical distance measure

In order to demonstrate the functionality of the *cd* and *qd* measures, 2D Pareto fronts (PTV coverage vs. Rectum dose) as well as sub-optimal treatment plans were created for five cases (one front per case) of prostate cancer, and *cd* and *qd* values were calculated between plans (see Supplementary appendix for plan generation details).

Though, the *qd* value was introduced above as a plan quality difference measure between fronts or between a front and a dominated plan, 2D and *n*D *qd* values were also calculated between plans on the front and used as part of a quality control with the purpose to verify that the treatment plans that defined the Pareto fronts truly belonged to the fronts. If the calculated 2D *qd* values between plans in the trade-off region of the evaluated parameters of a front were (almost) zero, the plans were considered to truly belong to the front (as per definition, the plan quality difference should be zero between such plans). If a 2D *qd* value was not zero, a variation was present in some of the plan evaluation parameters other than the two Pareto evaluated ones (i.e. another trade-off influencing the optimization). If a 2D *qd* value was not zero but at least the *n*D *qd* value was, the plan was still considered as being optimal but not part of the 2D Pareto front. If the *n*D *qd* value was not zero, the plan was considered as being sub-optimal.

The calculated 2D *cd* and *n*D *qd* values between dominated plans and plans on the fronts were compared to the results of a clinical treatment plan evaluation (clinical grading analysis (CGA) study [12]) in order to determine if the measures were representative of a plan quality difference, as assessed by decision makers (radiation oncologists and medical physicists). For the CGA, a reference plan on the Pareto front (Pareto plan) as well as eight sub-optimal plans for each case were exported to Oncentra® (Elekta AB, Stockholm, Sweden) TPS. In Oncentra®, the blinded plans were shown in pairs (Pareto plan and one sub-optimal plan) and compared side-by-side. Dose-distributions in all CT-slices, dose-volume histograms (DVH), and dose statistics for all targets and OARs were shown. Twelve observers, six radiation oncologists (RO), and six medical physicists (MP) from our local radiotherapy treatment center participated in the CGA study. They were asked to compare and assess the quality difference between the treatment plans. Each of their assessments was represented by a global plan quality grade (Table A1 in Supplementary appendix). The grades from all observers were compiled to see for which comparisons the observers could agree that there was a plan quality difference between the plan on the Pareto front and the sub-optimal plan. Sign tests were performed ($\alpha = 0.05$) to test the statistical significance of the results. Finally, these results were compared to the 2D *cd* and *n*D *qd* measures, including a check of the (Pearson product-moment) correlation between the 2D *cd* value for a plan and the frequency of D and D & E grades it was given, respectively, by the observers during the CGA.

Robustness tests were performed on the scaling values used (Table A2 in Supplementary appendix), i.e. to test how sensitive the results were to any change in scaling values, or in other words, to define for what choices of values the results were valid. During the test, the scaling values were varied, the *n*D *qd* values were calculated and compared against the CGA results. In this way, the range of scaling values could be found for which the measure was in agreement with the CGA results.

3. Results

The calculated distance measure and the results from the CGA were in a good agreement. The *cd* values correlated strongly ($r = 0.74$ and 0.90 for 27 and 21 points, respectively) with the number of D & E grades and (only) E grades awarded to the plans by the observers during the clinical plan quality evaluation (Fig. 1 and A3, Tables A3 and A4). At a *cd* value of around 0.32 (0.28–0.35 for all five cases), the observers in the CGA agreed ($p < .05$) that there was a difference in plan quality between the Pareto optimal and the sub-optimal plans. At $cd = 0.85$, the Pareto optimal plans were considered as much better than the sub-optimal plans ($p < .05$). The corresponding *n*D *qd* threshold value where the observers agreed upon a plan quality difference was at a lower and more precise value of 0.23 for all cases (dashed lines in Tables A3 and A4). The robustness tests performed on the values of the