Physics and Imaging in Radiation Oncology 5 (2018) 37–43

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

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

Original Research Article

Comparison of complexity metrics for multi-institutional evaluations of treatment plans in radiotherapy

Victor Hernandez[a,1,*], Jordi Saez[b,1], Marlies Pasler[c], Diego Jurado-Bruggeman[d], Nuria Jornet[e]

[a] Department of Medical Physics, Hospital Universitari Sant Joan de Reus, IISPV, Tarragona, Spain
[b] Department of Radiation Oncology, Hospital Clínic de Barcelona, Spain
[c] Lake Constance Radiation Oncology Center Singen-Friedrichshafen, Germany
[d] Medical Physics and Radiation Protection Department, Institut Català d'Oncologia, Girona, Spain
[e] Servei de Radiofísica i Radioprotecció, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

ARTICLE INFO

Keywords:
Plan complexity
Complexity metrics
Audits
Clinical trials

ABSTRACT

Background and purpose: It is known that intensity-modulated radiotherapy plans that are highly complex might be less accurate in dose calculation and treatment delivery. Multiple complexity metrics have been proposed, but the relationships between them have not been thoroughly investigated. This study investigated these relationships in multi-institutional comparisons of treatment plans, where plans from multiple treatment planning systems (TPSs) are typically evaluated.

Materials and methods: A program was developed to compute several complexity indices and provide analysis of dynamic plan parameters. This in-house software was used to analyse plans from a recent multi-institutional audit. Additionally, 100 clinical volumetric modulated arc therapy (VMAT) plans from two institutions using different TPSs were analysed.

Results: All plans produced satisfactory pre-treatment verification results and, hence, complexity metrics could not be used to predict plans failing QA. Regarding the relationship among complexity indices, some very strong correlations were found ($r > 0.9$ with $p < 0.01$). However, some relevant discrepancies between complexity indices were obtained, even with negative correlation coefficients ($r \sim -0.6$) which were expected to be positive. These discrepancies could be explained because each complexity index focused on different features of the plan and different TPSs prioritised modulation of different plan parameters.

Conclusions: Some complexity indices provided similar information and can be considered equivalent. However, indices that focused on different plan parameters yielded different results and it was unclear which complexity index should be used. Careful consideration should be given to the use of complexity metrics in multi-institutional studies.

1. Introduction

Advances in the technology for planning and delivery of radiotherapy treatments allow for highly conformal dose distributions to be achieved. However, these distributions require modulation of many machine parameters [1–5]. Since additional sources of variability are thus introduced, treatment plans with similar dose distributions may differ greatly in their complexity. Many investigators have reported that the degree of plan complexity may affect the accuracy of dose calculations and treatment delivery [6–13], which is crucial in dosimetry audits and clinical trials, as well as for big data analysis [14–16]. Therefore, aspects such as quality and complexity of treatment plans have to be carefully evaluated in multi-institutional plan comparisons [17].

Several investigators have proposed different complexity metrics and have reported correlations with overall accuracy and the resulting quality assurance (QA) metrics [6–13]. Thus, less complex plans offer several benefits such as more accurate dose calculations, more accurate and robust treatment delivery, better QA metrics and even lower risk of intra-fraction movements and patient variations [6–10]. For all these reasons plans with low complexity are associated with lower uncertainties and can be considered, in general, more robust than highly complex plans.

AAPM pointed out the need to incorporate measures of beam modulation to ensure that centres achieve intensity-modulated radiation-therapy (IMRT) plans that are comparable with regards to their

* Corresponding author.
E-mail address: vhernandezmasgrau@gmail.com (V. Hernandez).
[1] Both authors equally contributed to this study.

https://doi.org/10.1016/j.phro.2018.02.002
Received 30 September 2017; Received in revised form 30 January 2018; Accepted 1 February 2018

complexity [18]. However, it is not clear which of the proposed complexity indices should be used and the relationship between these multiple indices in multi-TPS environments has not been previously addressed. In this study we investigated the use of complexity metrics in multi-institutional comparisons where multiple TPSs, planners and linac types are typically involved. The study focused on volumetric modulated arc therapy (VMAT) treatments, but most of the indices evaluated can also be applied to other techniques such as sliding window and step-and-shoot IMRT.

2. Materials and methods

2.1. Complexity metrics

In this study several complexity indices that are computed from the treatment plan parameters defined at each of the control points of the plan were investigated. These indices allow for a detailed analysis of the dynamic parameters involved in treatment plans, which makes them more appropriate for VMAT than 'fluence-based' indices. The following indices were evaluated:

a) Modulation Complexity Score (MCS) [6]. This score integrates two contributions to complexity: variability in the shape of segments and variations in their area. MCS uses a fixed range from 0 to 1 and, unlike the rest of the complexity indices, it is defined in such a way that the lower the value of the MCS the higher the complexity. It was initially designed for step-and-shoot treatments and later adapted to sliding window and VMAT [9,19].
b) Edge metric (EM) [8]. This metric computes the complexity of multileaf collimator (MLC) apertures based on the ratio of MLC side edge length and aperture area. In this study the original recommendation for the parameters (C1 = 0 and C2 = 1) was followed. Thus, the greater the differences between the positions of adjacent leaves the higher the EM index, which is closely related to the amount of tongue-and-groove effect.
c) Leaf travel (LT) [9]. This index indicates the average distance travelled by the moving leaves. LT was devised for VMAT treatments consisting of a single full arc. To allow for simple comparisons between plans with a different number of arcs or with partial arcs, we divided LT by the corresponding arc length (typically about 360 deg for single arcs and about 720 deg for double arcs) and we named this index as 'LT/AL'.
d) Plan irregularity (PI) and Plan modulation (PM) [11]. PI describes the deviations of aperture shapes from a circle, being 1 for a perfect circle. PM indicates to what extent a beam is modulated with multiple smaller segments.
e) Modulation index total (MItotal) [10]. This index evaluates the variations in speed and acceleration of the MLC as well as variations of the gantry speed and the dose rate. MItotal is, to our knowledge, the only complexity index that takes into account the modulation of the dose rate and the gantry speed.

2.2. Treatment plans evaluated

The first group of plans evaluated in this study included forty plans from a recent audit promoted by the Catalan Association of Medical Physics within the framework of the Catalan-Occitan Oncology Group (GOCO). This audit included local pre-treatment verification results and independent dosimetry audit measurements [20]. A mock head-and-neck and a mock prostate case adapted from those proposed in TG119 were used. Most plans (twenty-eight) were produced with Eclipse™ (Varian Medical Systems), eight plans were generated with Pinnacle Auto-Planning (Philips Radiation Oncology Systems) and four plans with Monaco (Elekta AB). Hereafter these TPSs will be called TPS-A, TPS-B and TPS-C, respectively. Details on the TPSs, the linacs and the methodology used can be found in the aforementioned publication.

Additionally, in the present study clinical plans from TPS-A and TPS-B were also analysed. In particular, fifty head-and-neck VMAT plans and fifty prostate VMAT plans from each TPS were randomly selected and evaluated. Plans from TPS-A and TPS-B were produced for a Varian Clinac iX (Millennium 120 MLC) and an Elekta Synergy (MLCi2, binned dose rate), respectively.

2.3. Software and equipment used

To compute the previously described complexity indices, an in-house program called PlanAnalyser was developed in MATLAB (Mathworks, Massachusetts, USA). This software reads the DICOM plan as exported from the TPS and computes complexity indices using the data contained in the DICOM plan. Plan complexity indices were computed by joining all beams and performing the calculations for the 'combined' beam.

PlanAnalyser incorporates an emulator that predicts the variations of the dynamic plan parameters during treatment delivery. Since one of the complexity metrics (MItotal) evaluates the variations of the dose rate and gantry speed, we investigated the modulation of these parameters. Mean variations were defined as the total variation (i.e., sum of all variations between consecutive control points) divided by the total arc length. To verify the predictions from the emulator they were compared to results from log files for both Elekta and Varian linacs. Varian log files were analysed with in-house software [21] and log files from Elekta were recorded with the service graphing module of the linac controller (Integrity 1.2).

Pre-treatment verifications were carried out for all plans. Audit plans were measured with both independent QA equipment (ArcCHECK, Sun Nuclear Corporation) and a large variety of local QA devices [20]. Clinical plans from TPS-A and TPS-B were measured with ArcCHECK and Octavius II – 2D array seven29 (PTW Freiburg), respectively. Since audit plans corresponded to the same mock cases, a plan quality score was computed with the software PlanIQ™ (Sun Nuclear Corporation) in order to identify which plans achieved the best trade-off between target coverage, homogeneity, conformity, and doses to organs at risk [20].

To investigate the dependencies among these indices, the Spearman's rank correlation coefficients *r*, sensitive to both linear and non-linear correlations, were calculated. The strength of the association, for absolute values of r, 0–0.19 was regarded as 'no correlation', 0.20–0.39 as 'weak', 0.40–0.59 as 'moderate', 0.60–0.79 as 'strong' and 0.80–1 as 'very strong'. To account for multiple testing, false discovery rates (q-values) [22,23] were calculated. Reported p-values represent statistical analysis without multiple testing correction and statistical significance was considered at $p < 0.05$ with q-value < 0.1. All statistical analysis was performed in R-3.3.2 (R: A Language and Environment for Statistical Computing, 2016, Vienna, Austria).

3. Results

3.1. Audit plans

All participating centres fulfilled all the requested planning goals regarding both target coverage requirements and dose limits to organs at risk. Large differences in the degrees of plan complexity were observed, but no statistically significant correlation was found between dosimetric plan complexity and plan quality [20]. Pre-treatment verification results were clinically acceptable for all plans (> 95% of points with gamma 3%/3 mm < 1), hence complexity metrics could not be used to predict plans failing QA.

Regarding the comparison between complexity indices, strong correlations were found between MCS, PI and EM. However, we also observed some evident discrepancies, meaning that some plans were more complex than others according to a particular complexity index, while the opposite result was found when another complexity index was