Use of plan quality degradation to evaluate tradeoffs in delivery efficiency and clinical plan metrics arising from IMRT optimizer and sequencer compromises

Joel R. Wilkie, Martha M. Matuszak, Mary Feng, and Jean M. Moran
Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan 48109

Benedick A. Fraass[a)]
Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan 48109 and Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, California 90048

(Received 19 September 2012; revised 9 May 2013; accepted for publication 12 May 2013; published 10 June 2013)

Purpose: Plan degradation resulting from compromises made to enhance delivery efficiency is an important consideration for intensity modulated radiation therapy (IMRT) treatment plans. IMRT optimization and/or multileaf collimator (MLC) sequencing schemes can be modified to generate more efficient treatment delivery, but the effect those modifications have on plan quality is often difficult to quantify. In this work, the authors present a method for quantitative assessment of overall plan quality degradation due to tradeoffs between delivery efficiency and treatment plan quality, illustrated using comparisons between plans developed allowing different numbers of intensity levels in IMRT optimization and/or MLC sequencing for static segmental MLC IMRT plans.
Methods: A plan quality degradation method to evaluate delivery efficiency and plan quality tradeoffs was developed and used to assess planning for 14 prostate and 12 head and neck patients treated with static IMRT. Plan quality was evaluated using a physician's predetermined "quality degradation" factors for relevant clinical plan metrics associated with the plan optimization strategy. Delivery efficiency and plan quality were assessed for a range of optimization and sequencing limitations. The "optimal" (baseline) plan for each case was derived using a clinical cost function with an unlimited number of intensity levels. These plans were sequenced with a clinical MLC leaf sequencer which uses >100 segments, assuring delivered intensities to be within 1% of the optimized intensity pattern. Each patient's optimal plan was also sequenced limiting the number of intensity levels (20, 10, and 5), and then separately optimized with these same numbers of intensity levels. Delivery time was measured for all plans, and direct evaluation of the tradeoffs between delivery time and plan degradation was performed.
Results: When considering tradeoffs, the optimal number of intensity levels depends on the treatment site and on the stage in the process at which the levels are limited. The cost of improved delivery efficiency, in terms of plan quality degradation, increased as the number of intensity levels in the sequencer or optimizer decreased. The degradation was more substantial for the head and neck cases relative to the prostate cases, particularly when fewer than 20 intensity levels were used. Plan quality degradation was less severe when the number of intensity levels was limited in the optimizer rather than the sequencer.
Conclusions: Analysis of plan quality degradation allows for a quantitative assessment of the compromises in clinical plan quality as delivery efficiency is improved, in order to determine the optimal delivery settings. The technique is based on physician-determined quality degradation factors and can be extended to other clinical situations where investigation of various tradeoffs is warranted. © *2013 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4808118]

Key words: IMRT, optimization, plan scoring

I. INTRODUCTION

Delivery efficiency is an important consideration in intensity modulated radiation therapy (IMRT). The increased complexity of IMRT plans leads to delivery times that can often be at least twice as long as those for conventional plans.[1] Delivery efficiency depends both on the optimization and multileaf collimator (MLC) sequencing algorithms used in the treatment planning process, as well as on the specific method of IMRT delivery employed. Various more efficient optimization and sequencing strategies, as well as direct aperture optimization (DAO) and direct machine parameter optimization (DMPO) techniques, have previously been investigated.[2–15] These are designed to reduce the number of monitor units (MUs) or segments in a treatment plan, with the aim of reducing overall time and/or making the plans more robust to motion. Efficiency comparisons for different methods of static gantry IMRT delivery have also been previously reported.[16–19] The focus on delivery efficiency has intensified with the recent increased interest in rotational treatment delivery techniques

such as volumetric modulated arc therapy (VMAT).[14,20–22] These techniques aim to reduce treatment time by delivering radiation while the beam is rotating, and investigators have reported on improved delivery efficiency for VMAT relative to static gantry IMRT for several treatment sites.[23–29]

In general, more efficient delivery methods tend to reduce the available degrees of freedom in the plan optimization. This reduction can potentially degrade the plan which is finally delivered through compromises made during MLC sequencing or elsewhere in the planning process due to a reduction in the number of delivery degrees of freedom (e.g., limitations in intensity levels, number of segments, minimum size of segments, etc.). As a result, the final plan quality may be inferior to that of the optimal plan. However, the extent of plan quality degradation is often unknown, since the dose distribution from the optimal plan is not usually made available to the planner for review. Moreover, the effect of compromises made in plan quality on treatment outcome is not easily measured. The tradeoff in delivery efficiency and plan quality is, therefore, difficult to assess, and the ideal operating point on the quality versus efficiency curve cannot be readily determined. However, a method to clinically score plan quality degradation itself could be used to assess the tradeoff with delivery efficiency.

Quantitative evaluation and comparison of plan quality, as well as differences in plan quality which are achieved with different IMRT optimization and sequencing methods, have been investigated.[18,19,30–32] These studies commonly compare values of target dose homogeneity, maximum or average critical structure doses, or a penalty score from the cost function used in the optimization system. However, multiple plan quality factors are difficult to combine into an overall measure of clinical plan quality without direct physician input. Interpretation of the penalty cost is often complicated by the different strategies which may be employed by different treatment planners. In particular, expansions applied to contours of targets and avoidance structures may differ, along with the weights assigned to each objective in the overall cost function. Details of cost function implementations can also differ. Achieving plan comparison scoring which incorporates all the clinical concerns of the physician on a given case is even more difficult, and all of these issues together make quantitative comparison of different plans challenging.

In this work, we illustrate a method for evaluation of the quality versus delivery efficiency tradeoffs which are often imbedded in many IMRT and VMAT planning processes. This method does not attempt to do overall plan quality scoring, rather, it is intended only to characterize the clinical tradeoffs which are made in an individual case as plan quality is compromised to improve delivery efficiency. In this work we demonstrate the technique by comparing the plan degradation associated with intensity level quantization (typically applied in IMRT plan sequencing or indirectly in DAO or DMPO through the choice of allowable segments). In order to provide a quantitative measure of overall plan quality degradation, we employ physician-determined quality degradation factors (QDFs) that relate all relevant plan metrics. Using this method, clinical plan quality can be assessed quantitatively and the tradeoff with delivery efficiency can be more thoroughly investigated.

Specifically, we use this method to evaluate the tradeoff between delivery efficiency and clinical plan quality as the number of intensity levels used in segmental MLC (SMLC) delivery is varied for static gantry IMRT. The number of intensity levels used in SMLC delivery correlates with the number of segments needed to deliver the intensity profile for each treatment field.[19,30] Treatment time increases as more segments are delivered due to the increased overall time for MLC movement between segments, in addition to any per segment treatment control system overhead that may exist, and may increase due to use of more monitor units if smaller segments are used to create the final field. However, using more intensity levels in the optimizer may allow for a more optimal plan to be achieved, and using more intensity levels in the sequencer may prevent further degradation of the optimal plan. Previous investigators found that clinically acceptable plans could be achieved with as few as five intensity levels.[19,30–32] However, this may not be the optimal number in terms of the tradeoff between delivery efficiency and plan quality. Even if a plan is of acceptable quality, it may be possible to improve its quality without sacrificing much in terms of delivery efficiency. An understanding of how much plan quality degradation occurs, for a given method of increasing delivery efficiency, makes it possible to determine the optimal range for the number of intensity levels. In this research, we use the quality degradation factor method to investigate this tradeoff for a series of clinical prostate and head and neck cases.

II. MATERIALS AND METHODS

II.A. Cases

Fourteen consecutive prostate cases and 12 consecutive head and neck cases that were treated with IMRT in our department and met the prescription dose criteria were selected for this retrospective analysis. The prescription dose for the prostate cases was 77.7 Gy in 1.85 Gy fractions to the planning target volume (PTV), which consisted of the prostate and some or all of the seminal vesicles plus a 5 mm expansion. Organs at risk for these cases included the rectum, bladder, penile bulb, and femoral heads. The prescription dose for the head and neck cases was 70 Gy in 35 fractions for the high dose PTV, while mid-dose PTV prescriptions were 59 or 63 Gy, and low dose PTV prescriptions were 56 or 59 Gy. Organs at risk for these cases included the spinal cord, brainstem, oral cavity, mandible, lips, larynx, parotids, submandibular glands, esophagus, cochlea, brachial plexus, and pharyngeal constrictors.

Treatment planning for all plans was performed with our in-house-developed 3D planning system (UMPlan)[33] and inverse planning system (UMOpt).[34,35] All cases were planned for a Varian 21 EX linear accelerator with a 120 leaf multileaf collimator (MLC). 6 MV beamlet IMRT was used for planning the head and neck cases and 16 MV beamlet IMRT (our clinical standard) was used for the prostate cases. All plans used 0.5 × 0.5 cm beamlets and the same beam directions which had been used clinically: prostate plans all had seven

fields, and the head/neck cases had nine fields (except for one case each with eight and seven fields). The plans were optimized with a simulated annealing-based optimization algorithm using dose and dose-volume-based objectives created with the previously described costlet functionality of the optimization system.[34] The optimized plans were sequenced using an in-house-developed and previously described[36] leaf sequencer based on the Bortfeld method.[37] All prostate cases were optimized using one specific cost function. All head/neck cases were optimized with the same cost function except for differing mid and low dose target dose costlets as required by the clinical prescription. One planner performed the planning for all cases, and one physician evaluated all plans. All optimizations were run until they converged. All plans, including the baseline optimal plans (not limited in the number of intensity levels), have been performed using the standard clinical tools used within the department so all can successfully pass the clinical QA criterion [using 1 mm gradient-compensation,[38] all detectors in the multi-ion chamber IMRT QA device with dose $>80\%$ of the plan maximum and with <10 cGy spread (across the individual ion chamber) should be within $\pm 4\%$, a much tighter criterion than the often-used gamma of 3% and 3 mm for a composite delivery at planned gantry angles]. This was confirmed with IMRT QA measurements for an example prostate and example head and neck for a subset of plans at multiple intensity levels utilizing our 2D ionization chamber array (Scanditronix MatriXX) and in-house developed analysis tools.

II.B. Treatment plan optimization and sequencing

The optimized plan from the clinically employed cost function was used as the starting point for each case and was termed the optimal or baseline plan. This plan was optimized using an unlimited number of intensity levels in order to maximize the degrees of freedom available to the optimizer so it could achieve the optimal solution. The optimal plan was sequenced with the clinical MLC sequencer,[36] which uses 1% intensity level steps and up to 250 segments per beam, in order to assure delivered intensities are within 1% of the optimized intensity pattern.

For the study comparisons, the compromises made by the MLC sequencer and/or optimization system to limit the number of intensity levels (or segments) were separately evaluated. To evaluate sequencer-related compromises, the optimal plan was resequenced three additional times, using 20, 10, and 5 intensity levels ("Seq20," "Seq10," "Seq5," respectively). To evaluate making the same intensity level compromises, except within the optimizer, the optimal plan was further optimized three times, with 20, 10, and 5 intensity levels available in the optimizer ("Opt20," "Opt10," "Opt5," respectively). For these optimizations, the beamlet intensities from the optimal plan were initially rounded to the nearest fixed intensity level, and then the optimization was restarted using the same cost function and run until the penalty cost converged. The newly optimized plans were then sequenced using the same number of intensity levels in the sequencer as were used in the optimizer.

II.C. Delivery efficiency and quality degradation factor

Sequencer and optimizer-related compromises were evaluated by assessing the tradeoffs between delivery time and clinical plan quality for the reduced intensity level plans compared to the optimal plan. Total beam delivery time was determined by delivering each plan on a Varian 21iX equipped with a Millennium MLC and summing the beam-on time for each beam.

Clinical plan quality was assessed using physician-assigned clinically relevant plan metrics. These metrics were determined before the experiment and then used as a means to quantify the degradation in clinical plan quality for the reduced intensity level plans relative to the optimal plan. Plan evaluation metrics included D_{99} (the dose covering 99% of the highest dose volume), D_1 (the mean dose to the maximum 1% of the volume), V_{75} (volume of the 75% isodose surface), D_{mean} (the mean dose), and others.

To have a systematic measure of plan degradation for each site, a physician assigned the relative importance of degradation that might occur in each normal tissue metric for each target and critical structure. The goal of these assignments was for the physician to describe the relative importance of change in each metric as it affects the physician's plan evaluation. The overall quality degradation score is then directly relatable to changes in the clinical metrics they have chosen to evaluate when comparing rival plans. Defining a method which gives a single overall quality degradation score is very important, since plan quality compromises usually affect both target coverage and normal tissue doses, and comparison of dose volume histograms alone is unable to quantify the varying amounts of improvement and/or degradation that occur to the various structures with a given technique change. The QDF for a structure can be defined by pairwise comparison of the importance of a change in one metric with respect to another. Typically we have defined QDFs with respect to changes in the primary target metric, using a QDF of 1 for a 1 Gy decrease in the target dose metric value. For example, if a 1 Gy decrease in the minimum target dose is just as important as a 10 Gy increase in the mean dose metric for a normal tissue, then the QDF for the normal tissue would be 0.1/Gy, since $0.1/\text{Gy} \times 10\ \text{Gy}$ gives the same value as the 1 Gy change in target dose.

For this work, the QDFs for the prostate and head/neck cases were defined by a single physician before the experiment was performed. For all the experiments, a QDF of 1.0 was assigned to a 1 Gy decrease in the main target metric, D_{99} (the dose covering 99% of the highest dose PTV volume). All other QDFs were assigned by the physician based on the relative importance of changes in each metric according to the physician's clinical judgment, using comparison of each metric (one by one) to the 1 Gy decrease in the target metric (as described above). For a given plan, each relevant metric change from that achieved by the optimal plan was multiplied by its corresponding QDF, and the sum of those products was called the plan quality degradation (QD). The QD is a relative measure of degradation for plans evaluated according to the chosen metrics. The comparisons of QDs for two different

TABLE I. Physician-assigned QDFs for prostate cases per Gy or % as indicated in the table. A QDF of 1 represents a 1 Gy decrease in the D_{99} coverage of the PTV.

Structure	PTV		Rectum			Bladder			Bulb	Femurs
Parameter	D_{99}	D_1	D_1	V_{70}	V_{50}	D_1	V_{75}	V_{50}	Mean	D_1
QDF	1	1	1	0.2	0.2	0.5	0.1	0.1	0.1	1 if >45 Gy
Units	Gy	Gy	Gy	%	%	Gy	%	%	Gy	Gy

sites or for very different evaluation schemes are not expected to be directly informative, since the QD is only intended to quantify plan degradation for a given site, set of clinical metrics, and physician evaluation scheme.

Physician-assigned QDFs are listed in Table I for the prostate cases. This table illustrates that deviations for the rectum were considered more important for plan quality, *by this particular physician*, than those for the bladder. For head and neck cases, a QDF of 1.0 was assigned to the primary target metric, D_{99}, for the 70 Gy PTV, and smaller QDFs were assigned for coverage of the lower dose PTVs (Table II). The cord structures were defined to be the most important normal structures, followed by the brachial plexus and cochlea. Note that the QDF does not define the importance of the structure as a whole, it only defines the importance of changes to that particular metric for that structure. It is of course possible to evaluate and include multiple dosimetric (or other) metrics for a given structure, each with their own QDF.

For each case, the beam-on time was measured for the optimal plan and all six reduced intensity level plans. The plan quality degradation relative to the optimal plan was also calculated for these six plans. The clinical plan quality versus beam delivery time tradeoff was assessed for each case.

III. RESULTS

III.A. Prostate cases

The isodose distribution from an optimal plan for a typical prostate case (optimized with an unlimited number of intensity levels in the optimizer) is shown in Fig. 1(a) to illustrate the quality of the optimal plans. The plan is highly conformal with the 77 Gy isodose line nearly completely covering the PTV, little dose above 83 Gy, and steep dose gradients in the direction of the rectum and bladder. Figures 1(b)–1(d) show the dose difference as the number of intensity levels used in the sequencer is reduced. Figures 1(e)–1(g) show the dose differences which result when the plan is reoptimized (in the optimizer) using smaller numbers of intensity levels. In both cases, dose differences from the optimal plan become more substantial as fewer intensity levels are used, as one would expect. The maximum differences in both cases are typically less than 2 Gy for 20 and 10 intensity levels but exceed 3–4 Gy when only 5 intensity levels are used. However, when the changes are made in the sequencer alone, the dose differences seemingly are randomly spread around, leading to a larger decrease in the PTV dose homogeneity [as evidenced by the increased hot and cold regions in Fig. 1(d)]. This may also lead to increased maximum doses in organs at risk if random beamlet increases occur in the hottest regions.

Clinical DVH criteria and beam-on time for this example case are listed in Table III. Data for the femoral heads and penile bulb are not listed as they remained essentially unchanged regardless of the number of intensity levels used. The data indicate that most of the plan quality degradation comes from the PTV D_{99} and D_1 values (due to their high priorities in the cost function and importance to the physician's evaluation), with bigger deviations for sequencer-reduced intensity levels. The total plan quality degradation for each method of reducing intensity levels is plotted versus beam-on times in Fig. 2. A quality degradation of 1 is equivalent to a 1 Gy decrease in D_{99} for the PTV. The plot demonstrates that the tradeoff

TABLE II. Physician-assigned QDFs for head and neck cases. A QDF of 1 represents a 1 Gy decrease in the D_{99} coverage of the high dose (70 Gy) PTV. Note that for serial critical structures, the penalty was only applied if the D_1 value exceeded the planned dose limit for the structure.

Structure	PTVhigh		PTVmid		PTVlow		
Parameter	D_{99}	D_1	D_{99}	D_1	D_{99}	D_1	
QDF	1	1	0.5	1	0.3	1	
Units	Gy	Gy	Gy	Gy	Gy	Gy	
Structure	Parotids	Oral cavity	Larynx	Esophagus	Submandibular glands	Lips	Constrictors
Parameter	D_{mean}	D_{mean}	D_{mean}	D_{mean}	D_{mean}	D_{mean}	D_{mean}
QDF	0.1	0.1	0.2	0.1	0.1	0.1	0.1
Units	Gy	Gy	Gy	Gy	Gy	Gy	Gy
Structure	Cord	Cord + 5	Mandible	Brachial plexus	Cochlea		
Parameter	$D_1 > 45$ Gy	$D_1 > 50$ Gy	$D_1 > 70$ Gy	$D_1 > 60$ Gy	$D_1 > 40$ Gy		
QDF	10	10	1	3	3		
Units	Gy	Gy	Gy	Gy	Gy		

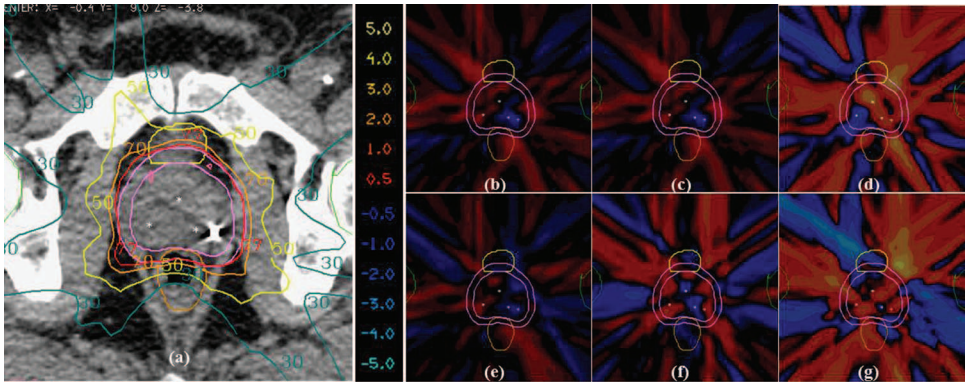

FIG. 1. Isodose distributions from the optimal plan for a prostate case (a) along with dose difference distributions (b)–(h) between reduced intensity level plans and the optimal plan. The number of intensity levels used was reduced to 20, 10, and 5 in the sequencer (b)–(d) and in the optimizer (e)–(g). Dose differences are shown in absolute dose (Gy).

in plan quality degradation versus beam-on time is relatively small in going from an unlimited number of intensities to 20 intensity levels, but that the degradation then becomes larger with each successive reduction in number of intensity levels. This graph can be interpreted as illustrating that choosing a plan which has been limited to only five intensity levels in the MLC sequencer has been degraded (in total) as much as decreasing the minimum dose to D_{99} by 4 Gy, while if that limitation of intensity levels had been made in the optimizer, the total degradation would have been halved, to only the equivalent of a 2 Gy compromise in D_{99}. Perhaps most interestingly, these compromises may be unknown to the planner if the system is always used with the five intensity level choice.

Plan quality degradation versus beam-on time for all 14 prostate cases is plotted in Fig. 3. In all cases, plan quality degradation increases in the same way as the number of intensity levels decreases, and more or less degrades approximately the same amount as the number of intensity levels is decreased.

III.B. Head and neck cases

The isodose distribution from the optimal plan for a typical head and neck case is shown in Fig. 4. Plan complexity for the head and neck cases is clearly greater than that for prostate cases. The minimum dose target coverage for all three PTVs is adequate, but the hotspots in the PTV70 target are larger than seen for most prostate cases. The dose difference plots in Fig. 4 for the Opt20, Opt10, Opt5 and Seq20, Seq10, Seq5 plans indicate that it is also more difficult to maintain the same plan quality as the number of intensity levels is reduced, particularly when they are reduced in the sequencer. Maximum target doses and PTV coverage noticeably suffer when ten or five intensity levels are used in the sequencer. Due to the often seemingly random nature of beamlet intensity modifications, the dose to some critical structures increases while the dose to others decreases. This is shown in Table IV, which lists the pertinent dose information as well as the beam-on times for all plans. The changes are more controlled when the number of intensity levels is reduced in the optimizer, although plan quality clearly suffers with this method as well.

A plot of plan quality degradation versus beam on time for this example case is shown in Fig. 5. A plan quality degradation of 1.0 is equivalent to a decrease in D_{99} for PTV70 of 1 Gy. It is again evident that plan quality degradation is more substantial when reducing the number of intensity levels in the sequencer compared to reducing it in the optimizer. However, it is also clear that the degradation values are larger

TABLE III. Clinical DVH criteria for the plans shown in Fig. 1. The values for the optimal plan are shown, while the differences from that optimal plan metric are shown for the sequencer (Seq) and optimization (Opt) defined intensity level plans.

Structure	Criteria	Optimal value	Seq 20	Seq 10	Seq 5	Opt 20	Opt 10	Opt 5
PTV	D_{99}	77.3	0	−0.5	−1.6	−0.1	−0.2	−0.7
	D_1	83.3	+0.2	+0.2	+1.8	0	+0.1	+0.7
Rectum	D_1	77.9	+0.1	+0.3	+0.3	0	0	+0.5
	V_{70}	11.9	0	+0.1	0	0	+0.1	−0.3
	V_{50}	19.9	0	+0.1	0	0	+0.1	−0.1
Bladder	D_1	74.5	−0.1	+0.1	+0.2	−0.1	+0.1	+0.6
	V_{75}	2.8	0	0	0	0	0	+0.1
	V_{50}	6.6	0	0	+0.1	0	0	+0.1
Time	Minutes	5.00	−1.32	−1.77	−2.20	−1.38	−1.78	−2.23

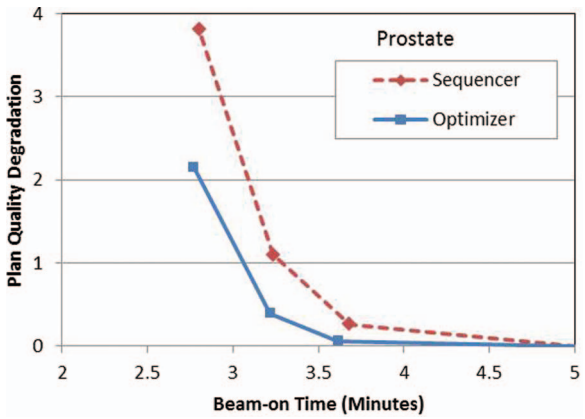

FIG. 2. Plan quality degradation versus beam-on time for a sample prostate case as the number of intensity levels is reduced in either the sequencer (dashed red line) or the optimizer (solid blue line). Labels on the graph show the number of intensity levels used.

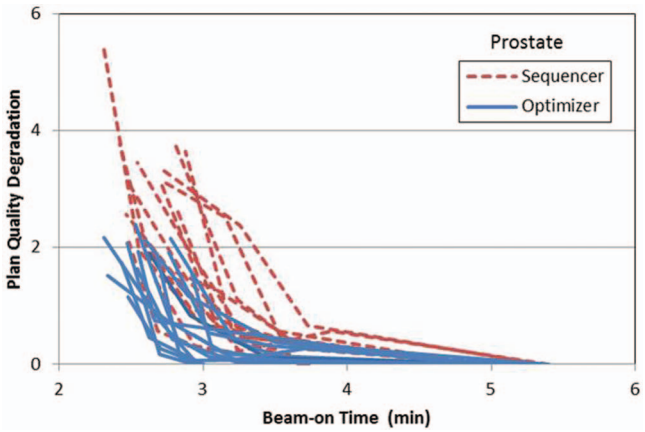

FIG. 3. Plan quality degradation versus beam-on time for all prostate cases as the number of intensity levels is reduced in either the sequencer (dashed red lines) or the optimizer (solid blue lines).

for this case compared to the prostate case. The beam-on times are also greater, meaning more time savings are possible relative to the optimal plan. This is also shown in Fig. 6, where the plan quality degradation is plotted versus beam-on time for all 12 cases. For each case, the tradeoff in plan quality when reducing from unlimited intensity levels to 20 levels is relatively small, but the degradation increases substantially when the number of intensity levels is changed from 20 to 10, and even more from 10 levels to 5.

III.C. Overall analysis

The average tradeoff (and standard deviation, shown by the error bars) for each successive intensity level reduction in the sequencer and in the optimizer, for both prostate and head/neck cases, is plotted in Fig. 7. For both prostate situations, the average tradeoff in going from an unlimited number of intensity levels in the optimizer down to 20 levels degrades the plan less than 0.1 Gy (in D_{99}) (per minute of decreased delivery time) for prostate and 0.2 Gy (in D_{99} of PTV70)/(min saved) for the H/N cases. As the number of intensity levels is further decreased from 20 to 10 and then from 10 to 5, the prostate tradeoff QDF averages go from 1 to 5.9 Gy (in D_{99} per minute saved) for prostate, and from 1.8 to 10.5 Gy/(min saved) for the H/N cases. The quality degradation tradeoffs are substantially larger when the number of intensity levels is reduced in the sequencing step alone, especially for the first two intensity level reductions. This last decrease to five intensity levels degrades prostate plans by a total degradation which is equal to underdosing the dosimetric prescription by 5.2 Gy, on average, a substantial degradation (according to the evaluation scheme developed for these cases). For the H/N cases, cutting the number of intensities from ten to five can cause quality degradation of as much as 15 Gy/min saved. Though there are differences that depend on the type of case, the quality degradation per minute saved is within a factor of 2 for the two very different types of cases, so it is possible this parameter is reasonably consistent and independent of the type of case.

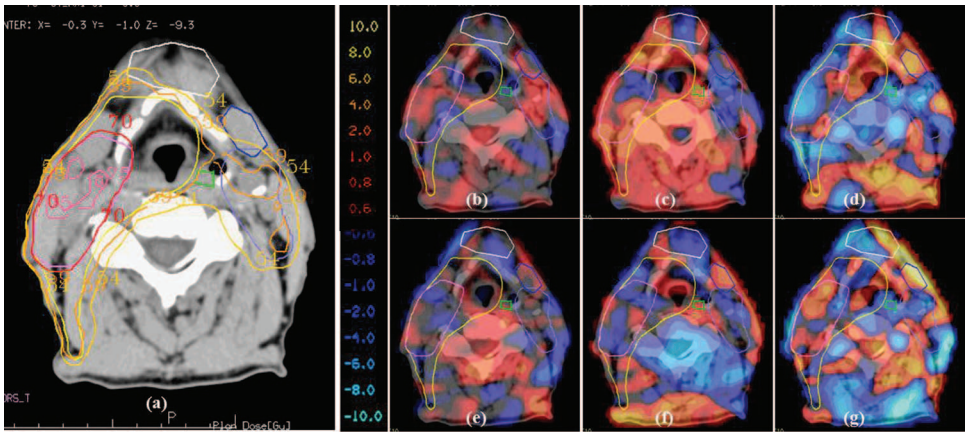

FIG. 4. Isodose distributions from optimal plan (a) for a head and neck case along with dose difference distributions (b)–(h) between reduced intensity level plans and the optimal plan. The intensity levels were reduced to 20, 10, and 5 in the sequencer (b)–(d) and in the optimizer (e)–(g). Dose differences are shown in absolute dose (Gy).

TABLE IV. Clinical DVH criteria for the plans shown in Fig. 4. The serial critical structures are not listed because the dose limits were not exceeded for any of the plans. The value for the optimal plan is shown, while the differences from that optimal plan metric are shown for the sequencer (Seq) and optimization (Opt) defined intensity level plans.

Structure	Criteria	Optimal value	Seq 20	Seq 10	Seq 5	Opt 20	Opt 10	Opt 5
PTVhigh	D_{99}	68.2	−0.3	−1.5	−4.4	−0.1	−1.0	−2.5
	D_1	74.5	0.6	2.8	4.5	0.1	0.8	2.6
PTVmid	D_{99}	57.7	−0.4	−1.6	−3.1	−0.2	−0.7	−2.0
	D_1	70.5	+0.7	+2.6	+3.6	+0.2	+0.7	+1.6
PTVlow	D_{99}	54.6	−0.5	−1.8	−4.2	−0.1	−0.6	−2.3
	D_1	59.8	+0.9	+1.4	+5.5	+0.2	+0.5	+2.6
Right parotid	D_{mean}	36	+0.4	−1.4	−1.7	+0.1	+0.8	+0.9
Left parotid	D_{mean}	21.3	−0.1	+0.1	−0.8	+0.7	−0.2	+2.0
Oral cavity	D_{mean}	29.9	−0.1	+0.1	+0.1	−0.2	+1.0	+2.4
Esophagus	D_{mean}	20.2	+0.2	0	−1.1	+0.2	+0.1	+0.9
Left submandibular	D_{mean}	45.5	+0.1	−1.1	+3.1	+0.1	+1.5	+2.7
Constrictors	D_{mean}	44.6	+0.1	−2.2	−0.5	−0.1	+0.6	+2.7
Time	Minutes	9.90	−4.80	−6.15	−6.94	−4.82	−5.89	−6.72

IV. DISCUSSION

In this work the tradeoff between clinical plan quality and delivery efficiency has been assessed for IMRT prostate and head and neck plans using the plan quality degradation method. The method has been demonstrated using example comparisons for step and shoot IMRT delivery based on the use of different numbers of intensity levels in the sequencer or the optimizer.

The fact that plans degrade as the number of intensity levels used is decreased has been demonstrated by many authors, as has the fact that plans become less complex and faster to deliver as the number of intensity levels decrease. In this paper, however, the plan quality degradation is defined by a single clearly defined clinically relevant metric tuned to the specific goals of a physician for that patient, so that the compromises made with various optimization and delivery decisions have a specific clinical importance. For example, for the data in Fig. 7, choosing to change the number of intensity levels used in the sequencer for a head/neck case from optimal to 10 will cause a plan degradation value of 4/min of delivery time: saving 1 min of delivery time will result in plan degradation equivalent to a decrease of 4 Gy in the minimum dose to the target volume. The physician then decides if saving 1 min of delivery time is worth a minimum target dose value 4 Gy lower than the desired value. Rarely have plan efficiency delivery compromises been quantified so clearly to the physician who must make the tradeoff decision.

Plan quality degradation, relative to the optimal plan, was much more substantial (by a factor of 2 or more) when the number of intensity levels was reduced in the sequencer than when the optimizer was used with that same limited number of intensity levels. This is not surprising, since some of the loss of plan quality can be avoided when optimization based on the plan's cost function is applied with the remaining intensity levels. The plan quality degradation was also more substantial for head and neck cases than for prostate cases with the same reductions in number of intensity levels. This is also expected, since the head and neck plans were more complex, with more PTV volumes and critical structures. However, the beam-on times were longer for the head and neck cases, so

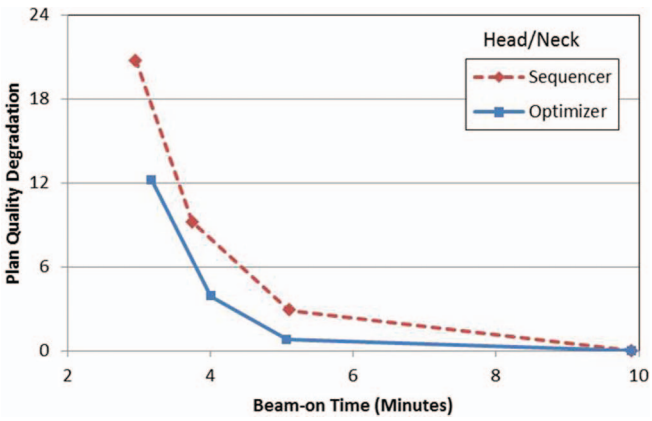

FIG. 5. Plan quality degradation versus beam-on time for sample head and neck case as the number of intensity levels is reduced in either the sequencer (dashed red line) or the optimizer (solid blue line).

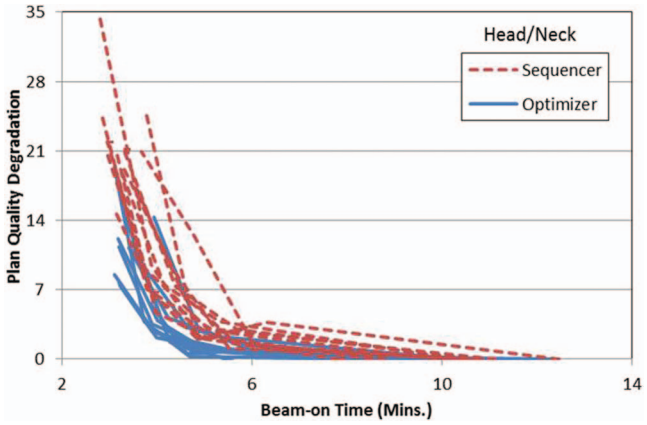

FIG. 6. Plan quality degradation versus beam-on time for all head and neck cases as the number of intensity levels is reduced in either the sequencer (dashed red lines) or the optimizer (solid blue line).

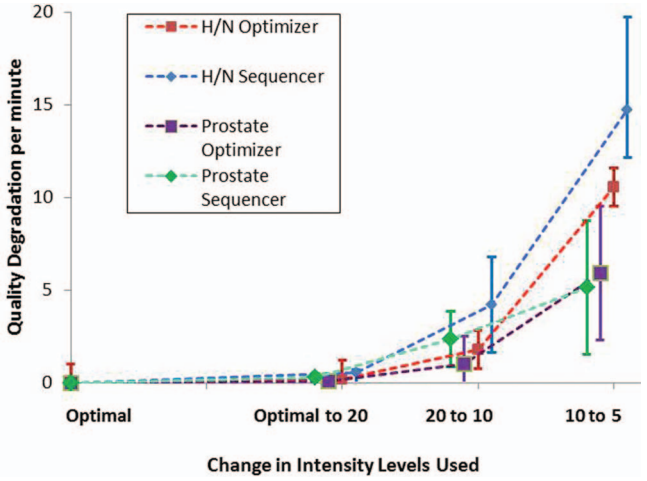

FIG. 7. Average additional plan quality degradations for each successive decrease in number of intensity levels (in the sequencer and the optimizer) for prostate and head and neck cases are shown in terms of degradation per minute. Error bars correspond to the standard deviations.

greater time savings (relative to the optimal plans) could be achieved with the head and neck plans when the number of intensity levels was reduced. Therefore, tradeoffs in plan quality versus delivery efficiency were more similar for the two treatment sites than one might initially expect, and may in fact be close enough that the quality degradation per minute saved may be a relatively consistent parameter for different types of cases.

The tradeoff plots can be used as a guide to determine the optimal number of intensity levels for these IMRT cases. This requires a determination of the acceptable plan quality degradation (per minute saved) for a particular treatment site. For example, if the acceptable cutoff was determined to be 2 QDF/min (i.e., we are willing to allow a decrease in target minimum coverage of 2 Gy/min of delivery time saved), then the ideal operating point on the tradeoff curve would be approximately 10 intensity levels in the optimizer for the prostate cases and closer to 20 intensity levels for the head and neck cases. All but one of the prostate cases meet this cutoff for 10 intensity levels while only 7 of the 12 head/neck cases meet this cutoff for 10 intensity levels. The optimal number of intensity levels in the sequencer alone would be 20 for both sites.

Understanding the kinds of plan degradation that occur due to standard planning system decisions may be important in many circumstances. For example, if an institution's standard planning technique were to limit the number of intensity levels to five, then the planner and physician may never know that the planning/optimization system has this degradation incorporated into its plans, and that the plans could potentially be significantly better with different parameter choices for the number of intensity levels used.

The plan quality degradation method described here can be used with any planning system and/or optimization method and/or changing planning technique, since it depends only on the clinical metrics chosen for plan evaluation and on the physician's choices about comparative importance of the various metrics. Each physician may in fact have their own set of QDFs for a given type of case, since the goal of the QDF choices is to quantify the physician's decisions about importance of various metrics as they evaluate plans. This technique is used here simply to help quantify the delivery efficiency versus plan quality tradeoffs, rather than the more general and difficult comparisons between very different types of plans. Given the limited goal for the quality degradation (i.e., determining the appropriate delivery efficiency versus plan quality tradeoff for a given case or cases), using this method should not introduce additional errors in the plan evaluation process, since it does not replace the clinical evaluation process, it simply helps the physician quantify the specific efficiency versus quality tradeoffs which might be available.

The physician-assigned QDFs (units of plan quality degradation) used in this study were considered to be linear for the sake of simplicity, though other behavior is certainly possible. For example, small changes in degradation may vary linearly with importance, but degradation that passes some threshold may become much more important (i.e., different than linear). For example, the QDF for a decrease in D_{99} of the PTV may get much larger (>1 QDF/Gy) when this value falls below 95% of the prescription dose. An increase in the mean parotid dose of 1 Gy may also have a greater impact on plan quality when the dose increases above 26 Gy. If these step-wise increases in plan quality degradation were incorporated into the analysis, the overall plan quality degradations would typically increase, particularly for more modulated plans. This would impact the degradation versus efficiency tradeoffs and could affect the determined optimal number of intensity levels.

It is important to note that the efficiency versus plan quality tradeoff depends on the machine and the specific type of IMRT delivery being used. Our results come from a modified Bortfeld sliding window SMLC delivery using a Varian 2100 iX linear accelerator with a 120 leaf MLC. For this machine and delivery method, the increase in delivery time with increasing number of segments is due almost entirely to leaf travel time between successive segments, as there is very little additional per segment overhead, and in these results there was little variation in the number of MUs with changes in the number of intensity levels. For delivery systems that add additional overhead for each segment, the tradeoff curve would be different, leading to a different choice of the optimal tradeoff between treatment time and plan quality.

The plan quality versus delivery efficiency tradeoff also depends on the treatment modality. The cost of improved plan quality, in terms of beam delivery time, depends on the degrees of freedom available and the plan complexity. For 3D-CRT treatments, the beam-on delivery time component comes only from the number of MUs, which can increase with the addition of subfields. The ideal number of fields is likely to be higher for more complex treatment sites where adding fields may have a more substantial impact on plan quality. For IMRT treatments, the number of degrees of freedom is typically large, and there may be the potential to decrease treatment times without substantial degradation of plan quality, particularly for less complex treatment geometries. On the other hand, improvements in plan quality may also be possible

without greatly increasing delivery time for IMRT plans. Plan quality for DAO and DMPO plans is dependent on the number of segments allowed as well as other details of the algorithms and implementation. For VMAT plans, the delivery time depends on gantry speed, dose rate, MLC leaf speed and distances to be traveled, and number of MUs. These factors are determined by plan complexity, and efficiency improvements may or may not substantially degrade plan quality. The quality versus delivery time tradeoff for VMAT (or versus fixed field IMRT) is another prime candidate for evaluation using the plan degradation method described in this work, since the method does not depend on any details of the optimization, and it only relates the various clinical metrics with which the plan is evaluated. Though some VMAT implementations have a very limited number of user-adjustable parameters, the same kind of quality compromises which occur in IMRT sequencing are often incorporated into the VMAT methods. Further study of VMAT-related tradeoff issues is the subject of ongoing research.

The key for any treatment modality is to determine the position of the current operating point on the degradation versus efficiency tradeoff curve in relation to the ideal operating point. This is also true for making comparisons between different treatment techniques. Without detailed study (for a given technique, or for comparisons of different techniques), this position on the curve may not be known. The clinically defined plan quality degradation metrics can be used for many techniques, and with care across different techniques, to help delineate the quality versus delivery efficiency tradeoffs and to determine the optimal techniques and parameters to be used for planning and delivery for a given treatment site and protocol.

V. CONCLUSIONS

This work has investigated the use of the plan quality degradation method to evaluate the tradeoff between plan quality and treatment delivery efficiency using example experiments with sets of prostate and head and neck IMRT cases using the SMLC delivery technique with varying numbers of intensity levels. The method allows for a quantitative assessment of the compromises in clinical plan quality as delivery efficiency is changed, in order to determine the optimal delivery settings. The plan quality degradation method is based on physician-assigned plan quality degradation factors and does not depend on the details of the plan optimization system used, so it may easily be extended to other treatment modalities, sites, and comparisons.

ACKNOWLEDGMENTS

This work was supported in part by NIH Grant No. P01-CA-59827. The authors thank Dr. Dan McShan and Dr. Dale Litzenberg for work on the optimization and sequencing software to make this study easier to perform.

[a]Author to whom correspondence should be addressed. Electronic mail: Benedick.Fraass@cshs.org; Telephone: 310-423-6133.

[1]G. A. Ezzell, J. M. Galvin, D. Low, J. R. Palta, I. Rosen, M. B. Sharpe, P. Xia, Y. Xiao, L. Xing, and C. X. Yu, "Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee," Med. Phys. **30**, 2089–2115 (2003).

[2]M. M. Matuszak, E. W. Larsen, and B. A. Fraass, "Reduction of IMRT beam complexity through the use of beam modulation penalties in the objective function," Med. Phys. **34**, 507–520 (2007).

[3]M. M. Coselmon, J. M. Moran, J. Radawski, and B. A. Fraass, "Improving IMRT delivery efficiency by applying intensity limits during inverse planning," Med. Phys. **32**, 1234–1245 (2005).

[4]X. Sun and P. Xia, "A new smoothing procedure to reduce delivery segments for static MLC-based IMRT planning," Med. Phys. **31**, 1158–1165 (2004).

[5]S. Spirou, N. Fournier-Bidoz, J. Yang, C. Chui, and C. Ling, "Smoothing intensity-modulated beam profiles to improve the efficiency of delivery," Med. Phys. **28**, 2105–2112 (2001).

[6]M. M. Matuszak, E. W. Larsen, K. Jee, D. L. McShan, and B. A. Fraass, "Adaptive diffusion smoothing: A diffusion-based method to reduce IMRT field complexity," Med. Phys. **35**, 1532–1546 (2008).

[7]J. L. Bedford and S. Webb, "Constrained segment shapes in direct aperture optimization for step-and-shoot IMRT," Med. Phys. **33**, 944–958 (2006).

[8]J. Dai and Y. Zhu, "Minimizing the number of segments in a delivery sequence for intensity modulated radiation therapy with a multileaf collimator," Med. Phys. **28**, 2113–2120 (2001).

[9]R. A. Siochi, "Minimizing static intensity modulation delivery time using an intensity solid paradigm," Int. J. Radiat. Oncol., Biol., Phys. **43**, 671–680 (1999).

[10]S. M. Crooks, L. F. McAven, D. F. Robinson, and L. Xing, "Minimizing delivery time and monitor units in static IMRT by leaf-sequencing," Phys. Med. Biol. **47**, 3105–3116 (2002).

[11]D. Craft, P. Suss, and T. Bortfeld, "The tradeoff between treatment plan quality and required number of monitor units in intensity-modulated radiotherapy," Int. J. Radiat. Oncol., Biol., Phys. **67**, 1596–1605 (2007).

[12]F. Carlsson, "Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy," Med. Phys. **35**, 3828–3838 (2008).

[13]B. A. Fraass, J. M. Steers, M. M. Matuszak, and D. L. McShan, "Inverse-optimized 3-D conformal planning: Minimizing complexity while achieving equivalence with beamlet IMRT in multiple sites," Med. Phys. **39**, 3361–3374 (2012).

[14]M. A. Earl, D. M. Shepard, S. Naqvi, X. A. Li, and C. X. Yu, "Inverse planning for intensity-modulated arc therapy using direct aperture optimization," Phys. Med. Biol. **48**, 1075–89 (2003).

[15]E. E. Ahunbay, G.-P. Chen, S. Thatcher, P. A. Jursinic, J. White, K. Albano, and X. A. Li, "Aperture optimization–based intensity-modulated radiotherapy for whole breast irradiation," Int. J. Radiat. Oncol., Biol., Phys. **67**, 1248–1258 (2007).

[16]D. Verellen, N. Linthout, G. Soete, S. Van Acker, P. De Roover, and G. Storme, "Considerations on treatment efficiency of different conformal radiation therapy techniques for prostate cancer," Radiother. Oncol. **63**, 27–36 (2002).

[17]X. R. Zhu, K. Prado, H. H. Liu, T. M. Guerrero, M. Jeter, Z. Liao, D. Rice, K. Forster, and C. W. Stevens, "Intensity-modulated radiation therapy for mesothelioma: Impact of multileaf collimator leaf width and pencil beam size on planning quality and delivery efficiency," Int. J. Radiat. Oncol., Biol., Phys. **62**, 1525–1534 (2005).

[18]P. Alaei, P. D. Higgins, R. Weaver, and N. Nguyen, "Comparison of dynamic and step-and-shoot intensity modulated radiation therapy planning and delivery," Med. Dosim. **29**, 1–6 (2004).

[19]C. Chui, M. F. Chan, E. Yorke, S. Spirou, and C. C. Ling, "Delivery of intensity-modulated radiation therapy with a conventional multileaf collimator: Comparison of dynamic and segmental methods," Med. Phys. **28**, 2441–2449 (2001).

[20]C. Cameron, "Sweeping-window arc therapy: An implementation of rotational IMRT with automatic beam-weight calculation," Phys. Med. Biol. **50**, 4317–36 (2005).

[21]K. Otto, "Volumetric modulated arc therapy: IMRT in a single gantry arc," Med. Phys. **35**, 310–317 (2008).

[22]C. X. Yu, "Intensity-modulated arc therapy with dynamic multileaf collimation: An alternative to tomotherapy," Phys. Med. Biol. **40**, 1435–49 (1995).

[23]A. Clivio, A. Fogliata, A. Franzetti-Pellanda, G. Nicolini, E. Vanetti, R. Wyttenback, and L. Cozzi, "Volumetric-modulated arc radiotherapy for carcinomas of the anal canal: A treatment planning comparison with fixed field IMRT," Radiother. Oncol. **92**, 118–124 (2009).
[24]L. Cozzi, K. A. Dinshaw, S. K. Shrivastava, U. Mahantshetty, R. Engineer, D. D. Deshpande, S. V. Jamema, E. Vanetti, A. Clivio, G. Nicolini, and A. Fogliata, "A treatment planning study comparing volumetric arc modulation with RapidArc and fixed field IMRT for cervix uteri radiotherapy," Radiother. Oncol. **89**, 180–91 (2008).
[25]A. Fogliata, A. Clivio, G. Nicolini, E. Vanetti, and L. Cozzi, "Intensity modulation with photons for benign intracranial tumours: A planning comparison of volumetric single arc, helical arc and fixed gantry techniques," Radiother. Oncol. **89**, 254–62 (2008).
[26]D. Palma, E. Vollans, K. James, S. Nakano, V. Moiseenko, R. Shaffer, M. McKenzie, J. Morris, and K. Otto, "Volumetric modulated arc therapy for delivery of prostate radiotherapy: Comparison with intensity modulated radiotherapy and three-dimensional conformal radiotherapy," Int. J. Radiat. Oncol., Biol., Phys. **72**, 996–1001 (2008).
[27]W. F. A. R. Verbakel, J. P. Cuijpers, D. Hoffmans, M. Bieker, B. J. Slotman, and S. Senan, "Volumetric intensity-modulated arc therapy vs. conventional IMRT in head-and-neck cancer: A comparative planning and dosimetric study," Int. J. Radiat. Oncol., Biol., Phys. **74**, 252–259 (2009).
[28]R. Shaffer, A. M. Nichol, E. Vollans, M. Fong, S. Nakano, V. Moiseenko, M. Schmuland, R. Ma, M. McKenzie, and K. Otto, "A comparison of volumetric modulated arc therapy and conventional intensity-modulated radiotherapy for frontal and temporal high-grade gliomas," Int. J. Radiat. Oncol., Biol., Phys. **76**, 1177–1184 (2010).
[29]M. Rao, W. Yang, F. Chen, K. Sheng, J. Ye, V. Mehta, D. Shepard, and D. Cao, "Comparison of Elekta VMAT with helical tomotherapy and fixed field IMRT: Plan quality, delivery efficiency and accuracy," Med. Phys. **37**, 1350–1359 (2010).
[30]M. Keller-Reichenbecher, T. Bortfeld, S. Levegrun, J. Stein, K. Preiser, and W. Schlegel, "Intensity modulation with the "step and shoot" technique using a commercial mlc: A planning study," Int. J. Radiat. Oncol., Biol., Phys. **45**, 1315–1324 (1999).
[31]X. Sun, P. Xia, and N. Yu, "Effects of the intensity levels and beam map resolutions on static IMRT plans," Med. Phys. **31**, 2402–2411 (2004).
[32]G. Nicolini, A. Fogliata, and L. Cozzi, "IMRT with the sliding window: Comparison of the static and dynamic methods. Dosimetric and spectral analysis," Radiother. Oncol. **75**, 112–119 (2005).
[33]B. A. Fraass and D. L. McShan, "3-D treatment planning: I. Overview of a clinical planning system," in *The Use of Computers in Radiation Therapy*, edited by I. A. D. Bruinvis, F. H. van der Giessen, H. J. van Kleffens, and F. W. Wittkamper (Elsevier, Amsterdam, 1987), pp. 273–276.
[34]M. L. Kessler, D. L. McShan, M. Epelman, K. A. Vineberg, A. Eisbruch, T. S. Lawrence, and B. A. Fraass, "Costlets: A generalized approach to cost functions for automated optimization," Optim. Eng. **6**, 421–448 (2005).
[35]J. H. Kim, N. Dogan, D. L. McShan, and M. L. Kessler, "An AVS-based system for optimization of conformal radiotherapy treatment plans," in *1995 International Advanced Visual Systems User and Developer Conference, Boston, MA* (Advanced Visual Systems, Boston MA, 1995), pp. 417–423.
[36]D. W. Litzenberg, J. M. Moran, and B. A. Fraass, "Incorporation of realistic delivery limitations into dynamic MLC treatment delivery," Med. Phys. **29**, 810–820 (2002).
[37]T. R. Bortfeld, D. L. Kahler, T. J. Waldron, and A. L. Boyer, "X-ray field compensation with multileaf collimators," Int. J. Radiat. Oncol., Biol., Phys. **28**, 723–730 (1994).
[38]J. M. Moran, J. Radawski, and B. A. Fraass, "A dose gradient analysis tool for IMRT QA," J. Appl. Clin. Med. Phys. **6**, 62–73 (2005).