Investigation of effective decision criteria for multiobjective optimization in IMRT

Clay Holdsworth,^{a)} Robert D. Stewart, Minsun Kim, Jay Liao, and Mark H. Phillips
Department of Radiation Oncology, University of Washington, Box 356043, Seattle, Washington 98195-6043

(Received 19 October 2010; revised 12 April 2011; accepted for publication 13 April 2011; published 27 May 2011)

Purpose: To investigate how using different sets of decision criteria impacts the quality of intensity modulated radiation therapy (IMRT) plans obtained by multiobjective optimization.

Methods: A multiobjective optimization evolutionary algorithm (MOEA) was used to produce sets of IMRT plans. The MOEA consisted of two interacting algorithms: (i) a deterministic inverse planning optimization of beamlet intensities that minimizes a weighted sum of quadratic penalty objectives to generate IMRT plans and (ii) an evolutionary algorithm that selects the superior IMRT plans using decision criteria and uses those plans to determine the new weights and penalty objectives of each new plan. Plans resulting from the deterministic algorithm were evaluated by the evolutionary algorithm using a set of decision criteria for both targets and organs at risk (OARs). Decision criteria used included variation in the target dose distribution, mean dose, maximum dose, generalized equivalent uniform dose (gEUD), an equivalent uniform dose ($EUD_{\alpha,\beta}$) formula derived from the linear-quadratic survival model, and points on dose volume histograms (DVHs). In order to quantitatively compare results from trials using different decision criteria, a neutral set of comparison metrics was used. For each set of decision criteria investigated, IMRT plans were calculated for four different cases: two simple prostate cases, one complex prostate Case, and one complex head and neck Case.

Results: When smaller numbers of decision criteria, more descriptive decision criteria, or less anti-correlated decision criteria were used to characterize plan quality during multiobjective optimization, dose to OARs and target dose variation were reduced in the final population of plans. Mean OAR dose and gEUD ($a = 4$) decision criteria were comparable. Using maximum dose decision criteria for OARs near targets resulted in inferior populations that focused solely on low target variance at the expense of high OAR dose. Target dose range, ($D_{max} - D_{min}$), decision criteria were found to be most effective for keeping targets uniform. Using target gEUD decision criteria resulted in much lower OAR doses but much higher target dose variation. $EUD_{\alpha,\beta}$ based decision criteria focused on a region of plan space that was a compromise between target and OAR objectives. None of these target decision criteria dominated plans using other criteria, but only focused on approaching a different area of the Pareto front.

Conclusions: The choice of decision criteria implemented in the MOEA had a significant impact on the region explored and the rate of convergence toward the Pareto front. When more decision criteria, anticorrelated decision criteria, or decision criteria with insufficient information were implemented, inferior populations are resulted. When more informative decision criteria were used, such as gEUD, $EUD_{\alpha,\beta}$, target dose range, and mean dose, MOEA optimizations focused on approaching different regions of the Pareto front, but did not dominate each other. Using simple OAR decision criteria and target $EUD_{\alpha,\beta}$ decision criteria demonstrated the potential to generate IMRT plans that significantly reduce dose to OARs while achieving the same or better tumor control when clinical requirements on target dose variance can be met or relaxed. © 2011 American Association of Physicists in Medicine. [DOI: 10.1118/1.3589128]

Key words: multiobjective, IMRT, optimization, radiation therapy

I. INTRODUCTION

A common approach to intensity modulated radiation therapy (IMRT) optimization is to minimize a predetermined weighted sum of *penalty objectives*. This method may produce suboptimal plans because the space of clinically feasible plans is not fully explored. In our earlier paper,¹ we described a multiobjective evolutionary algorithm (MOEA) that searches a larger region of plan space by varying both

parameters (penalty objectives and their weights) in the penalty function. In that context, proximity to the Pareto front as determined by an independent set of clinically relevant *decision criteria* was used to judge plan quality. In this work, the same algorithm is used to investigate how using different sets of decision criteria impacts the quality of the final set of IMRT plans produced. The goal is to investigate whether different decision criteria do a better job of searching regions of space relevant to clinical decision making.

In the conventional inverse planning approach, the beamlet intensities are optimized by minimizing a given penalty function which produces a treatment plan that is optimal only in the space of the penalty objectives used. The first optimization rarely produces a desirable plan, so typically the weights and the organ-specific penalty objectives are varied and the optimization is repeated. Our MOEA executes this process in a more complete and systematic manner and produces a set of optimal plans. In both planning approaches, the penalty objectives and weights are changed throughout the optimization process, and it becomes necessary to introduce some other method for comparing the results such as a radiation oncologist's judgment or a set of independent decision criteria. In our MOEA, a separate set of decision criteria is used in the multiobjective optimization, and the result is an approximation to the Pareto front in that space. While the mathematical form of penalty objectives are often chosen for ease of optimization, decision criteria can be tailored to be more clinically relevant.

Although the MOEA is a powerful tool for searching a wide region of possible plans without *a priori* information or preference, little is known about how physical and biological decision criteria impact selection and convergence to the Pareto front. In this paper, we compare the performance of different sets of decision criteria for use in multiobjective IMRT optimization. The comparisons are between sets of plans that are Pareto optimal in their respective decision criteria spaces. In order to make unbiased comparisons of various trials, plan quality is assessed using unbiased clinically relevant *comparison metrics* distinct from the decision criteria used to evolve optimal sets of IMRT plans.

II. METHODS AND MATERIALS

II.A. Methods of comparison

II.A.1. Pareto optimality

In the context of our MOEA, **penalty objectives** are used in a deterministic (inverse planning) optimization to create IMRT plans. The evolutionary algorithm judges plan quality and selects superior plans using a separate set of **decision criteria**. In comparing two plans with n decision criteria, A and B , the values of the criteria for plan A are determined by the vector $\mathbf{X}^A \in \mathbb{R}^n$. Plan A is said to dominate plan B if $X_i^A \leq X_i^B$ for all i with strict inequality for at least one coordinate. The Pareto front consists of a set of solutions that are not dominated by any other possible solution.

II.A.2. Independent metrics

It is difficult to compare the performance of various methods that use different decision criteria to judge IMRT plan performance. To accomplish this, we developed an independent set of **comparison metrics** (described in Sec. II.D) only for the purpose of using them as a tool for comparing methods in this work.

II.A.3. Domination comparison

When comparing two different strategies, the multicriteria aspect must still be taken into account.² Often, the ability of

an optimization algorithm to achieve certain types of objectives is correlated with its inability to achieve other objectives (this is termed as *anticorrelation*). A comparison of the methods must look at the results of both at the same time. For example, if one method achieved better target coverage in 50% of the plans relative to a second method and better normal tissue sparing in 50% of the cases, this could mean that half the time the method produced dominant plans, but it could also mean that half of the plans were superior in target coverage and inferior in tissue sparing and vice versa.

Domination comparison^{1,2} is a way to quantitatively compare competing sets of plans. Domination comparison, $D_C(A, B)$, is defined to be the percent of comparisons between plans from set A and set B that result in the plan from set A dominating (all comparison metrics being superior to) the plan from set B . $D_C(A, B)$ can also be thought of as the probability for a plan randomly selected from set A to dominate a plan randomly selected from set B . If $D_C(A, B) \gg D_C(B, A)$, it is likely set A approximates the true Pareto front better than set B . Domination comparison does not measure other qualities of multiobjective populations, such as diversity.

II.B. Optimization algorithm

Our MOEA was used to find sets of IMRT plans that approximate the Pareto front¹ utilizing an evolutionary algorithm coupled with a deterministic optimization algorithm. The evolutionary algorithm varies the parameters used in the penalty function, and then function is minimized using the deterministic algorithm, the MOEA selects the most fit plans as determined by decision criteria. The deterministic algorithm optimizes the set of beam intensities using a fast inverse planning algorithm in which a set of quadratic dose penalty objectives are weighted and summed, and the resulting single penalty function is minimized. Typically multiobjective optimizations only vary the weights and keep the penalty objectives constant, but the evolutionary algorithm is able to search different regions of plan space by providing both a new set of weighting factors and a new set of penalty objectives. The decision criteria are not used in the penalty function of the deterministic algorithm for a number of reasons, including the speed of optimization and possible limitations on the type of functions that can be used for optimization.

In general penalty functions are of the form

$$\sum_i \sum_j H_i(D_{ref,i}) W_i (D_{ref,i} - d_{i,j})^2 + F_S, \quad (1)$$

where $d_{i,j}$ is the calculated dose in the j^{th} voxel of the i^{th} structure, W_i the weight of the i^{th} structure, $H(D_{ref,i})$ is the Heaviside function that sets the penalty to zero in OAR voxels when $D_{ref,i} > d_{i,j}$ and is equal to 1 otherwise, and D_i is the reference dose parameter of the i^{th} structure. The smoothing function, F_S , is designed to make resulting beam configurations deliverable. The vectors, $\vec{W}$ and $\vec{D}_{ref}$, are genes within the MOEA, and each individual plan is characterized by a unique pair of vectors.

A commercial inverse planning system, Pinnacle³ (Philips Medical Systems) with the DMPO IMRT optimization, was also used in one comparison.

II.C. Decision criteria

It is rarely the Case, either clinically or in treatment planning comparison studies, that plans are selected solely on the basis of the penalty objectives. Typically, clinical judgment or a second set of decision criteria is used to decide on the “optimal” plan to be used for the treatment. Several sets of decision criteria were investigated for tumors and for organs at risk (OARs). The first set consisted of using the OAR penalty objectives as decision criteria. Additionally, this set of decision criteria also included the organ-specific reference doses, D_i , which were optimized by the MOEA. This set of decision criteria directly reflects the deterministic optimization process with the added goal of reducing the normal tissue reference doses as much as possible. The second set of OAR decision criteria investigated were dose volume histogram constraints often used by radiation oncologists.

Generalized equivalent uniform dose (gEUD) decision criteria that incorporate radiobiological information into the selection and ranking of plans (Refs. 2–5) were also investigated. The basic form of the function is

$$gEUD(\vec{d}) = \left(\frac{1}{nvox_i} \sum_{j=1}^{nvox_i} d_{i,j}^{a_i} \right)^{1/a_i}, \quad (2)$$

where $nvox_i$ = number of voxels in the i^{th} structure, $d_{i,j}$ is the dose in the j^{th} voxel of the i^{th} target, and a_i is a constant characteristic of the i^{th} structure. The effectiveness of using gEUD as decision criteria for the OARs with a set to 1.0, 4.0, or ∞ was investigated. The gEUD is equivalent to the mean dose when $a = 1$ and is equivalent to the maximum dose when $a = \infty$. Mean dose is commonly used as a dosimetric surrogate for complications in parallel tissues, and maximum dose is a commonly used dose surrogate for complications in serial tissues.

The alternate form of EUD ($EUD_{\alpha,\beta}$ used only for target decision criteria) incorporates a more complete model of cell killing and is described in Eq. (3).

$$EUD_{\alpha,\beta} = 0.5\alpha / (G * \beta) * (\sqrt{1 - (4\ln(S) / (\alpha * (\alpha/\beta)))} - 1), \quad (3)$$

where S is the fraction of surviving cells after the entire treatment, $G = 1/\text{number of fractions}$ (33 for this experiment), α is the linear component of tumor cell kill ($\alpha = 0.039$), and β is the quadratic component of tumor cell kill ($\beta = 0.026$).⁶

Typically, target dose distributions must meet minimum dose requirements as well as limits on inhomogeneity. No minimum dose decision criteria were required for any of the methods investigated, because dose distributions were either scaled to meet all target minimum dose criteria or were scaled to a fixed minimum $EUD_{\alpha,\beta}$ or gEUD. Three different methods of dose scaling and target decision objectives were evaluated. (a) All dose distributions were scaled so that all minimum dose target requirements were met, and the target

decision criteria was the dose range, $(D_{\max} - D_{\min})$; (b) dose distributions were scaled so that gEUD ($a = 0.25$) equaled or exceeded the required dose for all target volumes; and (c) the same as (b) but using $EUD_{\alpha,\beta}$ as described in Eq. (3).

For the latter two methods, no target decision criteria were used. In other words, the algorithm assumed that once the prescription dose is met, any extra dose to the target was not considered beneficial or detrimental.

II.D. Neutral comparison metrics for evaluating sets of decision criteria

Decision criteria determined the goals of the multiobjective optimization and the characteristics of the final solutions. To compare various strategies (optimizations using different sets of decision criteria), a neutral set of comparison metrics was needed to evaluate and rank plans in a way that avoids bias towards any specific decision criteria. Given the wide range of decision criteria that have been used, we chose two comparison metrics (one for OARs and one for targets) that are closely related to those used in clinical trials and not directly related to the sets of decision criteria being evaluated.

For targets, first all plans were normalized such that the amount of total tumor cell kill was equivalent to that of uniform minimum required dose. Then dose difference ($D_5 - D_{95}$), the least dose received by the top 5% of the volume minus the least dose received by the top 95% of the volume, was the comparison metric used to quantify dose variation in the target.

For an organ at risk, the OAR dose volume histograms (DVHs) for all plans in populations involved were averaged together to create a standard DVH. The OAR dose metric for a specific plan was then calculated by averaging the ratio of its DVH to the standard DVH for each integral dose from 1 Gy to $D(95\%)$. This comparison metric was normalized so that negative values corresponded to more optimal dose distributions lower than the standard DVH and positive values corresponded to less optimal dose distributions greater than the standard DVH. In other words, this metric measures the deviation from the “average” DVH. The 5% of the OAR volume that received the highest dose was not considered as ratios, involving such small volumes may become arbitrarily large.

Plots and domination comparison values with respect to the above set of comparison metrics were used to compare sets of IMRT plans generated using different sets of decision criteria in the MOEA.

II.E. Experiments

Two simple prostate example cases (Case 1 and Case 2), one complex prostate example Case (Case 3), and one head and neck example Case (Case 4) were considered. Prostate Case 1 and Case 2 had only three structures: PTV, rectum, and bladder. Prostate Case 1 was designed to be typical, and prostate Case 2 was designed to be more difficult, with more overlap between OARs and target. The target prescription was 70 Gy for both cases. The complex prostate, Case 3, used seven structures: GTV, PTV, rectum, bladder, left femur, right femur, and unspecified normal tissue, and the

FIG. 1. Results for first (left) and second (right) example prostate patients. Plot shows rectum and target comparison metrics for different strategies. The bladder comparison metric is reflected by the size of the symbols, larger symbols correspond to lower bladder dose. Results compare using gEUD for OAR decision criteria versus using dose parameter and corresponding penalty versus using four DVH points as decision criteria at V25, V50, V75, and V95.

nominal target prescription dose was 81 Gy. The head and neck Case, Case 4, was far more complex. One GTV and one PTV were used for the primary target with a prescription dose of 70 Gy. There were two more PTVs for secondary targets with prescription doses of 56.8 Gy and 62.7 Gy. In addition, there were ten separate OARs: right cochlea, left cochlea, right parotid, left parotid, brainstem, cord, epiglottis, submandibular, oral cavity, and unspecified brain tissue. For all PTVs (GTVs), the minimum allowed target dose was 95% (97%) of prescription dose, and only 5% (3%) of the target volume could be below the prescription dose when target range decision objectives were used. When target $EUD_{\alpha,\beta}$ (or gEUD) decision criteria were used, dose distributions were scaled so that the $EUD_{\alpha,\beta}$ (gEUD) for each target was at least equal to that target's required dose.

The following trials were performed to explore differences in the sets of plans obtained with alternate decision criteria:

- Comparison of common types of decision criteria for OARs: gEUD, the reference dose used and the corresponding penalty objective value, and four DVH points.
- Comparison of different levels of complexity in DVH decision criteria for OARs: one DVH point, two DVH points, and four DVH points.
- Comparison of different gEUD decision criteria for OARs: gEUD with $a_i = 1$ (mean dose), gEUD with $a_i = 4$, and gEUD with $a_i = \infty$ (maximum dose).
- Comparison of different target decision criteria and methods of dose scaling: target range ($D_{max} - D_{min}$) vs gEUD with $a_i = 0.25$ vs $EUD_{\alpha,\beta}$.

III. RESULTS

III.A. Comparison of common OAR decision criteria

Three different sets of decision criteria were used in the MOEA for all four example patients. These decision criteria

were chosen as representative of the types of functions used clinically to evaluate plans. They were also chosen to explore how the number of decision criteria and how anticorrelation amongst decision criteria impact results.

The first set included one decision criterion, gEUD ($a = 4$), for each OAR. The second set used two decision criteria: the reference dose parameter used in the penalty objectives, $D_{ref,i}$, and the penalty objective value, $\sum H(D_{ref,i} - d_{i,j})^2$, for each OAR, i.e., the penalty objectives were used as the decision criteria. The third optimization method used four decision criteria: the percent volume receiving at least 95% of the primary target's prescription dose (V95), V75, V50, and V25 for each OAR. All optimizations used dose variance ($D_{max} - D_{min}$) as the target decision criteria. After optimization and before decision criteria calculation, dose distributions were scaled such that all plans met minimum target dose requirements. In the MOEA, Case 1 optimized four genes (reference dose and weighting of penalty objectives for each of two OARs) and used three (gEUD), five (penalty), and nine (DVHs) decision criteria; Case 2 optimized four genes, and used three, five, and nine decision criteria; Case 3 optimized 11 genes, and used 7, 12, and 22 decision criteria; and Case 4 optimized 23 genes, and used 14, 24, and 44 decision criteria.

To compare sets of plans obtained using different decision criteria, unbiased comparison metrics were calculated using the DVHs of all final sets of plans. Lower comparison metric values correspond to more optimal dose distributions (lower OAR dose and reduced target variance). For each patient, the MOEA performed 300 inverse planning optimizations. Graphs comparing OAR and target comparison metrics of all three methods are shown for the two simple prostate patients in Figs. 1(a) and 1(b).

For both cases, the plans generated using OAR gEUD decision criteria tended to have lower target variance, lower rectum dose, and lower bladder dose. The other methods tended to have inferior plans, especially when the dose

Fig. 2. Results for complex prostate (left) and complex head and neck (right) example patients. The left plot shows rectum and target comparison metrics for different strategies, and the bladder objective is reflected by the size of the symbols, larger symbols correspond to lower bladder dose. The right plot shows the target comparison metric (x-axis), the averaged comparison metric values for all the OARs close to the target (y-axis), and the averaged comparison metric values for all the OARs far from the target (symbol size). Results compare using gEUD for OAR decision criteria versus using dose parameter and penalty objective values as decision criteria versus using four DVH points as decision criteria at V25, V50, V75, and V95.

parameter and penalty objectives were used as decision criteria. The same experiment was done for the two more complex cases. Since it is difficult to show more than three comparison metrics in a graph, the results for the complex prostate Case (Case 3) in Fig. 2(a) are shown for only the three most important structures: rectum, bladder, and PTV. For the complex head and neck Case, for purposes of display in Fig. 2(b), the comparison metrics for all four targets were averaged together (x-axis), the comparison metric for the four OARs closest to the targets were averaged (y-axis), and the comparison metric for five OARs furthest from the targets were averaged (symbol size is a function of the value of the objective).

The two complex cases have many more structures than the simpler cases. The corresponding greater number of decision criteria decreases the probability of one plan dominating another. In the MOEA, only dominated plans were eliminated. Even though a technique called domination advantage was used to enhance selective pressure, convergence was slower and much larger population sizes resulted when more decision criteria were used.¹ In the complex prostate Case, plans that used OAR gEUD decision criterion formed a Pareto front that tended to be superior to other plans, but this is more difficult to see than in the simple cases (see Table I). For the complex head and neck Case, each population had a large number of plans, the populations are not converged, and in order to display 14 comparison metrics in a single graph, multiple comparison metrics had to be averaged together. This makes interpretation of this figure challenging, but a large number of plans generated using two or four decision criteria are inferior to even the worst plan generated using gEUD OAR decision criteria. Table I shows relative domination comparison values with respect to comparison metrics between different methods for all four cases.

Table I reflects that the more comparison metrics used, the more difficult it is for every comparison metric of one plan to be superior to every comparison metric of another plan. For all four patients, using only one decision criterion, gEUD, for each OAR resulted in plans that were closer to the Pareto front than either of the other two methods. Using a greater number of decision criteria to describe the same information lowered the probability of domination among individual plans in the population, which weakened selective pressure. The selective pressure, and hence rate of convergence for the optimization, using penalty objectives and reference dose as decision criteria were further compromised because the paired objectives were anticorrelated, i.e., a lower values of $D_{ref,i}$ were likely to lead to higher values of the quadratic penalty objectives. Using two anticorrelated decision criteria proved to be more detrimental than using more decision criteria likely to have some degree of correlation.

III.B. Comparison of DVH points OAR decision criteria

Three different MOEA optimization strategies were considered for all four cases. The first strategy used one decision

TABLE I. Values of domination comparison between plans obtained using different sets of decision criteria. Each cell in the table represents the domination comparison values for one set versus the other and the reverse domination. For example in the top left cell, “gEUD plans” dominated “penalty plans” in 27.4% of possible comparisons, and 0% of penalty plans dominated gEUD plans.

	gEUD vs penalty	Penalty vs four DVHs	Four DVHs vs gEUD
Case 1	27.4% $\Rightarrow$ 0%	2.6% $\Rightarrow$ 10.2%	0% $\Rightarrow$ 28.1%
Case 2	13.3% $\Rightarrow$ 0%	0.3% $\Rightarrow$ 13.8%	0% $\Rightarrow$ 2.0%
Case 3	0.6% $\Rightarrow$ 0%	0.1% $\Rightarrow$ 0.5%	0% $\Rightarrow$ 0.2%
Case 4	0.05% $\Rightarrow$ 0%	0% $\Rightarrow$ 0%	0% $\Rightarrow$ 0.05%

Fig. 3. Results for complex prostate (left) and complex head and neck (right) example patients. The left plot shows rectum and target comparison metrics for different strategies. The bladder comparison metric is reflected by the size of the symbols, larger symbols correspond to lower bladder dose. The right plot shows the average target comparison metric (x axis), the averaged comparison metric values for all the OARs close to the target (y axis), and the averaged comparison metric values for all the OARs far from the target (symbol size). Results compare using one (V75), two (V95 and V50), or four (V25, V50, V75, and V95) DVH points as decision criteria.

criterion (V75) for each OAR. The second method used two decision criteria (V50 and V95) and the third optimization method used four decision criteria (V95, V75, V50, and V25) for each OAR. After optimization and before decision criteria calculation, dose distributions were scaled such that all plans met minimum target dose requirements. The number of genes and decision objectives were the same as in the previous comparison. All optimization methods used dose variance ($D_{max} - D_{min}$) as a target decision criteria. Graphs comparing OAR and target comparison metrics of all three methods are shown only for the two complex cases in Fig. 3.

In the complex prostate Case, plans that used either one or two DVH points as OAR decision criteria tended to be superior to the plans generated using four DVH points as decision criteria, though each tended to focus on approaching a different region of the Pareto front. For the complex head and neck Case, plans generated using one DVH point as decision criteria for each OAR were generally superior to both the two DVH decision criteria per OAR and the four DVH decision criteria per OAR. Table II shows relative domination comparison values with respect to comparison metrics between different methods for all four cases. As can be seen,

better performance for a smaller number of decision criteria did not hold for the simple cases.

It is apparent from the results that having a smaller number of decision criteria is not the only factor in MOEA performance. While a decision criterion such as *gEUD* contains much information about the dose distributions over an entire OAR, a single point on a DVH provides information of only a small portion of the dose distribution. For the simple cases, using only one DVH point did not provide enough information to determine which was the superior of two plans, and the benefit to performance from more information outweighed the cost of an increased number decision criteria. In the complex cases, each DVH point used per OAR added five (complex prostate) or ten (head and neck) decision criteria. This increase in the number of decision criteria appears to be more of a factor than using the insufficient information from a single DVH point. In determining a proper set of decision criteria, the benefits of using a small number of decision criteria has to be balanced with both using enough decision criteria to model the multiple goals of the treatment planning and providing sufficient information on each dose distribution to determine when a plan is superior.

TABLE II. Domination comparison values between IMRT plans obtained using different numbers of DVH points in the OAR objectives as decision criteria. Each cell in the table represents the domination comparison values for one set versus the other and the reverse domination. For example, in the top left cell, “one DVH” plans dominated “two DVH” plans in 4.1% of possible comparisons, and “two DVH” plans dominated “one DVH” plans in 12.1% of possible comparisons.

	One DVH vs two DVHs	One DVH vs four DVHs	Two DVHs vs four DVHs
Case 1	4.1% $\Rightarrow$ 12.1%	11.3% $\Rightarrow$ 7.0%	15.4% $\Rightarrow$ 1.8%
Case 2	1.0% $\Rightarrow$ 3.1%	1.1% $\Rightarrow$ 1.5%	1.1% $\Rightarrow$ 2.9%
Case 3	0.08% $\Rightarrow$ 0.25%	0.33% $\Rightarrow$ 0%	0.23% $\Rightarrow$ 0.06%
Case 4	0% $\Rightarrow$ 0%	0.01% $\Rightarrow$ 0%	0% $\Rightarrow$ 0%

FIG. 4. Results for first (left) and second (right) example prostate patients. Plot shows rectum and target comparison metrics for different methods. The bladder comparison metric is reflected by the size of the symbols, larger symbols correspond to lower bladder dose. Results compare using gEUD, maximum dose, or mean dose OAR decision criteria.

III.C. Comparison of using different exponents in gEUD OAR decision criteria

Comparisons were made using three different values of the exponent in the gEUD function for OAR decision criteria: $a = 1$ corresponding to the mean organ dose, $a = 4$, and $a = \infty$ corresponding to maximum organ dose. Dose range, $(D_{max} - D_{min})$, was used as the target decision criteria for all optimizations. IMRT MOEA optimizations that generated 300 individual plans were performed using each set of decision criteria for each example patient. MOEA optimizations using different sets of decision criteria were compared, and results for the two simple prostate cases are shown in Fig. 4.

In the MOEA, Case 1 optimized four genes (weight and penalty objective for each OAR) and used three decision criteria for all comparisons; Case 2 optimized four genes and used three decision criteria; Case 3 optimized 11 genes and used 7 decision criteria; and Case 4 optimized 23 genes and used 14 decision criteria.

In the simple prostate cases, optimizations using mean dose and gEUD OAR decision criteria had similar overall performance with the “gEUD” population shifted slightly into space with more uniform target distributions and the “mean” population focused slightly more on lower OAR dose. Optimizations that used maximum dose decision criteria focused on reducing target dose at the expense of high dose for OARs near the target. These sets of plans were inferior for the simple cases, but were not dominated in the complex cases. Table III shows relative domination comparison values between different methods for all four cases.

$EUD_{\alpha,\beta}$ using mean dose OAR decision criteria resulted in lower overall OAR dose for OARs adjacent to the target, but the multiobjective algorithm could make up for this with improved target distributions when gEUD ($a = 4$) OAR decision criteria were used. In simple cases, using maximum dose OAR objective resulted in plans that only focused on improving target distributions and were inferior to other

methods. In the complex cases, some diversity was lost when maximum dose was used, but the plan populations were not dominated by other methods.

III.D. Comparing decision criteria used for targets

Three different sets of decision criteria and dose scaling schemes were used to appropriately scale dose distributions and determine the quality of target coverage. The first method was to scale IMRT plans so 95% of all target volumes received at least the required dose and all target minimum doses $\geq 95\%$ of prescribed dose, and use target range, $D_{max} - D_{min}$, for target decision criteria. The second method was to scale dose distributions such that all target gEUDs [$a = 0.25$ in Eq. (2)] were greater than or equal to the required dose and not use target decision criteria. The third method was similar to the second except that $EUD_{\alpha,\beta}$ from Eq. (3) was used to scale dose distributions. The latter two methods operated on the assumption that once all targets receive the required dose, additional dose was not beneficial or detrimental to the treatment. For all methods, mean dose was used for OAR decision criteria. Results for all four patients are shown in Fig. 5.

Scaling by target gEUD decision objectives resulted in the lowest OAR doses and the highest target variance for all

TABLE III. Domination comparison values with respect to comparison metrics between plans obtained using OAR decision criteria with different exponents in the gEUD function. Each cell in the table represents the values of domination for one set versus the other and the reverse domination.

	$a = 1$ vs $a = 4$	$a = 4$ vs $a = \infty$	$a = \infty$ vs $a = 1$
Case 1	1.0% $\Rightarrow$ 1.0%	21.5% $\Rightarrow$ 0%	0% $\Rightarrow$ 16.9%
Case 2	0.7% $\Rightarrow$ 0%	10.9% $\Rightarrow$ 0%	0% $\Rightarrow$ 14.1%
Case 3	0% $\Rightarrow$ 0%	0% $\Rightarrow$ 0%	0% $\Rightarrow$ 0%
Case 4	0% $\Rightarrow$ 0%	0% $\Rightarrow$ 0%	0% $\Rightarrow$ 0%

FIG. 5. Results for case 1 (upper left), case 2 (upper right), case 3 (lower left), and case 4 (lower right). Plot shows rectum (averaged “close” OARs for case 4) and target comparison metrics for optimizations that scale dose distributions to meet minimum target dose requirements, to match all target $EUD_{\alpha,\beta}$ s to the required dose, or to match all target gEUDs to the required dose. The bladder (averaged “far” OARs) comparison metrics are reflected by the size of the symbols, larger symbols correspond to lower dose.

four cases. Scaling by minimum target requirements and using target range decision criteria resulted in the most uniform targets at the expense of high OAR doses. When dose distributions were scaled using target $EUD_{\alpha,\beta}$ decision criteria, plans were generated that represented a compromise between low OAR dose and target uniformity. None of these strategies generated populations of plans that were superior to any other strategy; they simply focused on different regions of plan space. In all four cases, not one IMRT plan dominated any other plan generated using any other set of decision criteria.

III.D.1. Target range versus target $EUD_{\alpha,\beta}$ versus clinical methods

For an example of clinical relevance, two 900-iteration IMRT MOEA optimizations were performed for a complex head and neck Case with ten OARs and four targets, and resulting dose volume histograms were compared directly.

One MOEA optimization used target dose range, $D_{max} - D_{min}$, as decision criteria and scaled dose distributions to meet all minimum target requirements. The second scaled dose distributions so that the $EUD_{\alpha,\beta}$ with $\alpha = 0.23$ and $\beta = 0.027$ (Refs. 7 and 8) of all targets were at least the minimum required dose. Nine hundred MOEA iterations had to be performed to get good convergence due to the greater number of decision criteria.

Both MOEA optimizations used OAR mean decision criteria and a maximum dose decision criteria for an “unspecified brain” OAR to reduce hot streaks through normal tissue. The most clinically desirable plan was chosen from each optimization strategy for comparison. Beams were translated into deliverable beams using a simple sweep technique,⁹ and a clinical plan was included in the comparison [Figs. 6(a)–6(d) and Table IV].

The IMRT plan generated using target range decision objectives had OAR and target dose distributions that were superior to the clinical plan. The plan generated by the

FIG. 6. Dose volume histograms of plans using three optimization methods with different methods of meeting target requirements: an MOEA optimization with target range decision criteria scaled by minimum dose requirements (*T-Range*); an MOEA optimization that used target $EUD_{\alpha,\beta}$ to scale plans without using target decision criteria (*T-EUD*); and a commercial optimization algorithm (*Clinic*).

MOEA optimization using target $EUD_{\alpha,\beta}$ to scale plans resulted in much lower dose in all OARs than both the other MOEA optimization and the clinical plan for the same or better biological effectiveness (number of cancer cells killed) of dose delivered to each target (see Table I). However, there are cold spots in the dose distributions of both secondary PTVs (see Fig. 6). Scaling by target $EUD_{\alpha,\beta}$ increased the flexibility of the MOEA optimizations, and the smaller number of decision objectives increased selective pressure and focused the search on reducing OAR dose. According to the biological model used to calculate the $EUD_{\alpha,\beta}$, this IMRT plan dominated both of the other plans; however, radiation oncologists may reject this plan because of the cold spots in both secondary PTVs.

IV. DISCUSSION

Clinical IMRT treatment planning typically collects a set of multiple penalty objectives into a single weighted penalty

function and finds plans optimal with respect to that set. Radiation oncologists then determine which IMRT plan is to be used in the clinic based on a number of clinical decision variables that often does not include the penalty objectives used for optimization.^{10–12} The decision criteria evaluated in this manuscript are designed to more directly model clinically relevant aspects of IMRT plan selection.

Our multiobjective optimization method is distinguished from other multiobjective approaches^{13,14} that generate plans that are Pareto optimal in the space of the inverse planning penalty objectives, and plans are not evaluated in a separate “decision criteria” space. Lahanas *et al.*¹³ mentions the need to do so, but the thrust of the paper is in establishing the basic Pareto fronts. Craft *et al.*¹⁴ have developed a user interface that allows the decision maker to apply whatever clinical criteria they desire upon inspection, but the set of plans generated are Pareto optimal only in penalty objective space. Our MOEA provides a set of plans that approximates the true Pareto front in the decision criteria space.¹ In addition, by optimizing both the form of the penalty objectives

TABLE IV. Head and neck case: $EUD_{\alpha,\beta}$ for all targets and the mean dose for all OARs are presented for three optimizations using different methods of meeting required target doses: clinical inverse planning, range of target dose, and target $EUD_{\alpha,\beta}$.

	Optimization method		
	Pinnacle ³ (Gy)	MOEA using target $D_{max} - D_{min}$ (Gy)	MOEA using target $EUD_{\alpha,\beta}$ (Gy)
Primary GTV	71.5	72.1	73.3
Primary PTV	70.1	72.0	73.0
PTV 62.7 Gy	65.5	65.3	65.4
PTV 57.0 Gy	58.8	59.0	58.7
Lt parotid	22.9	12.6	17.7
Rt parotid	39.6	39.7	23.7
Submandibular	22.9	7.2	4.9
Oral cavity	38.7	29.6	25.4
Epiglottis	45.4	33.6	5.6
Spinal cord	30.6	23.8	19.9
Brainstem	24.8	23.6	6.4
Rt cochlea	25.7	11.8	6.3
Lt cochlea	20.3	8.7	4.1

and their weights, our MOEA can search a larger space that is likely to include other clinically relevant plans.

The neutral comparison metric used to compare optimizations using different decision criteria was derived from the common clinical practice of inspecting the DVHs. The comparison metric attempts to fairly quantify the quality of all IMRT plans by their deviations from this average DVH.

In the first experiment, optimization performance as determined by comparison metrics using three different sets of decision criteria were compared. The combination of the penalty objectives and the OAR reference dose was chosen to represent the choice of the inverse planning objectives also being used as decision criteria which is common in comparisons of optimization results. Planners typically seek to reduce both the quadratic penalty value as well as the limiting OAR dose. The set of four DVH points was another effort to incorporate conventional clinical planning processes. Finally, $EUD_{\alpha,\beta}$ was chosen as being representative of biologically based decision criteria.

The paired penalty objective and dose parameter performed the worst. This is most likely because the two decision criteria are “anticorrelated” which makes finding plans that improve both decision criteria more difficult. These results do not mean that penalty objectives, if used by themselves, are poor decision criteria for methods that use fixed dose parameters in the penalty of objective functions. The optimization using $gEUD$ decision criteria did better than when four DVH points were used. The probability of domination decreases exponentially with a large number of objectives. This means using more decision criteria may result in slower convergence to the Pareto front, and it also means that domination comparison values reduced significantly as the cases got more complex and the number of comparison metrics increased.

The influence of the number of decision criteria was explored in the second experiment where we focused only

on DVH values at single points and varied the number of them. The results indicated that using a smaller number of decision criteria is not the only factor in finding superior plans. Enough information must be conveyed by a particular set of decision criteria to determine what dose distributions are superior. Of course, having enough criteria for the MOEA to consider distinct aspects of a multiobjective optimization is another important factor.

The superior performance of the $gEUD$ objectives led to a third experiment in which the exact form of the $gEUD$ function was varied for OAR decision criteria, since the best value of the exponent is not known. The $gEUD$ function at the extremes ($a = 1$ and $a = \infty$) are identical to the mean and maximum dose objectives which permitted some insight into the performance of these objectives. Using $a = 4$ or 1 in the $gEUD$ dose decision criteria gave similar results, with $gEUD_{a=4}$ exploring regions of space with slightly lower target variation and $gEUD_{a=1}$ exploring regions with slightly lower OAR dose. Using $gEUD_{a=\infty}$ decision criteria for OARs near targets resulted in populations that focused solely on low target dose variance at the expense of high OAR dose. In simple cases these populations were inferior to those obtained using $gEUD_{a=1}$ or $gEUD_{a=4}$, but in the complex cases, there was enough flexibility for other aspects of the multiobjective optimization to be improved, so populations were not dominated by other methods.

When evaluating different methods of scaling dose distributions and different sets of target decision criteria, it was found that no plan using any one method dominated any other plan using a different set of decision criteria for any patient case. Using target range ($D_{max} - D_{min}$) decision criteria was found to be most effective for keeping targets uniform, and using target $gEUD$ to scale dose distributions resulted in much lower OAR dose and higher target dose variance. Using target $EUD_{\alpha,\beta}$ to scale dose distributions, a region of plan space was explored that was a compromise between low target dose variance and low OAR dose. As judged by comparison metrics, each strategy explored a different region of space, but there was no advantage in convergence to the Pareto front.

Scaling the whole dose distribution to meet all target minimum dose requirements and then using target dose range, ($D_{max} - D_{min}$), as the decision criterion for targets resulted in the best performance when strict clinical limitations on target uniformity are enforced. Using $EUD_{\alpha,\beta}$ to scale plans without using target decision criteria, a biologically superior IMRT plan was found and target variance was reduced to a more acceptable level for most targets.

One of the advantages to using a multiobjective optimization algorithm is that the decision maker can be given a choice between plans that embody each of the goals to a different degree. Such an algorithm can also do a more thorough search for optimal parameters and plans, resulting in greater confidence that all aspects of the plans generated cannot be improved without degrading other aspects. However, regardless of the thoroughness of the overall search, the choice of decision criteria has a significant impact on the set of plans that are ultimately selected.

V. CONCLUSIONS

Using greater numbers of decision criteria to judge plan quality in IMRT MOEA tended to reduce the effectiveness of the selection method and worsen performance especially when such criteria were anticorrelated; however, enough decision criteria must be used to model distinct aspects of a multiobjective optimization and provide enough information to determine what plans are superior. When forms of decision criteria were used that provided sufficient information, such as $gEUD$, $EUD_{\alpha,\beta}(a_i = 4)$, and mean dose for OARs and targets, MOEA optimizations focused on approaching different regions of the Pareto front, but none of the populations were superior. Multiobjective IMRT optimization is a more powerful and versatile tool than single objective optimization. The use of a stochastic MOEA allows optimization with respect to more clinically relevant decision criteria than having to perform plan selection using penalty objectives that are designed to be easily optimized.

ACKNOWLEDGMENT

This work was supported by NIH Grant No. 1-R01-CA112505.

^{a)}Electronic mail: choldsw@u.washington.edu

¹C. Holdsworth, M. Kim, J. J. Liao, and M. H. Phillips, "A hierarchical approach to multiobjective optimization in IMRT," *Med. Phys.* **37**, 4986–4997 (2010).

- ²M. Phillips and C. Holdsworth, "When is better best? A multiobjective perspective," *Med. Phys.* **38**, 1635–1640 (2011).
- ³A. Niemierko, "A generalized concept of equivalent uniform dose (EUD)," *Med. Phys.* **26**, 1100 (1999).
- ⁴Q. Wu and R. Mohan, "Multiple local minima in IMRT optimization based on dose-volume criteria," *Med. Phys.* **29**, 1514–1527 (2002).
- ⁵Q. Wu, D. Djajaputra, H. H. Liu, L. Dong, R. Mohan, and X. Wu, "Dose sculpting with generalized equivalent uniform dose," *Med. Phys.* **32**, 1387–1396 (2005).
- ⁶J. Fowler, R. Chappell, and M. Ritter, "Is alpha/beta for prostate tumors really low?," *Int. J. Radio. Oncol. Biol. Phys.* **50**, 1021–1031 (2001).
- ⁷H. D. Thames, S. M. Bentzen, I. Turesson, M. Overgaard, and W. Van Den Bogaert, "Time-dose factors in radiotherapy: A review of the human data," *Radiother. Oncol.* **19**, 219–235 (1990).
- ⁸S. M. Bentzen and H. D. Thames, "Overall treatment time and tumor control dose for head and neck tumors: The dog leg revisited," *Radiother. Oncol.* **25**, 143–144 (1992).
- ⁹T. 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).
- ¹⁰V. A. Semenenko, B. Reitz, E. Day, X. S. Qi, M. Miften, X. A. Li, "Evaluation of a commercial biologically based IMRT treatment planning system," *Med. Phys.* **35**, 5851–5860 (2008).
- ¹¹X. S. Qi, V. A. Semenenko, and X. A. Li, "Improved critical structure sparing with biologically based IMRT optimization," *Med. Phys.* **36**, 1790–1799 (2009).
- ¹²D. N. Mihailidis, B. Plants, L. Farinash, M. Harmon, L. Whaley, P. Raja, P. Tomara, "Superiority of equivalent uniform dose (EUD)-based optimization for breast and chest wall," *Med. Dosim.* **35**, 67–76 (2010).
- ¹³M. Lahanas, E. Schreibmann, and D. Baltas, "Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms," *Phys. Med. Biol.* **48**, 2843–2871 (2003).
- ¹⁴D. L. Craft, T. F. Halabi, H. A. Shih, and T. R. Bortfeld, "Approximating convex Pareto surfaces in multiobjective radiotherapy planning," *Med. Phys.* **33**, 3399–3407 (2006).