Seminars in
RADIATION ONCOLOGY

Automated Radiotherapy Treatment Planning

Kevin L. Moore, PhD

The "treatment planning" component of managing a radiotherapy patient currently consumes hours, even days, of human effort. The time and workforce demands of the current planning paradigm can expose patients to delays and potentially substandard treatments, all while standing as seemingly insurmountable roadblocks to adaptive radiotherapy. Automating the treatment planning process is not a new idea, but recent advances have shown that automated planning might finally be turning the corner from niche research endeavor to a standard clinical practice. In this Seminar, we will examine the current state of automated radiotherapy planning, taking particular care to consider the most critical component of the planning problem: how to generate high-quality treatment solutions that account for patient-to-patient anatomical variation. Recommendations for testing and validating automated planning routines will be discussed, as well as potential drawbacks that might persist even after robust validation has been conducted.
Semin Radiat Oncol 29:209–218

Introduction

Before any consideration of automated planning in radiation therapy, it is worth lingering on the general concept of treatment planning itself. As an enterprise, "planning" has come to be known as the process by which a radiotherapy device is programmed to deliver an amount of radiation to a patient. For several decades, a radiotherapy "plan" could be described by a single value of "time" when a delivery device is energized or a radioisotope is in position. As the complexity of the delivery devices increased the parameter space of delivery techniques expanded in turn. Instead of just the duration of a teletherapy beam, the energy and shape of an external beam radiation field were now also under the clinician's control, and the specification of these became part of the plan for treating patients. As the field advanced, clinicians gained the ability to simulate both a patient's 3-dimensional (3D) anatomy and the dose deposition into that geometrical arrangement of tissues. With an understanding of where and how the radiation was being absorbed in a patient, the "quality" of the programmed machine instructions could be understood as an evaluation of the degree to which the simulated treatment maximized a therapeutic ratio of balancing tumor control against the complications associated with irradiating surrounding normal tissues. This process of (I) inputting a specific patient with disease X, (II) integrating current understanding of what a quality radiotherapy treatment is for disease X, (III) simulating the interaction of the patient and the treatment machine, and (IV) programming the treatment machine to maximize a therapeutic ratio comprises what has become known as the "planning" process that persists to this day.

Presently, the "Planning" step in the Figure 1 block diagram is a computer-aided but ultimately human-driven process that outputs a highly complex set of instructions to a delivery machine. That "planning" is computer-aided strikes no one as surprising, given that the simulated representation of the patient takes the form of digital medical images and the model of the delivery device exists as a computational algorithm. What perhaps should be surprising is that there is any human component at all to this process! While no one questions that they are highly skilled operators, human treatment planners could never analytically account for even the simplest 3D dose distributions in our patients, nor could they ever hope to program all but the most basic machine instructions that represent the primary output of their work. If these human operators cannot even parse the inputs and outputs of the system in which they operate, why do they still remain such an obligatory step in this process?

This question is akin to asking why, in 2019, there are still humans driving automobiles. Like the safety record of self-driving cars,[1] automated treatment planning has substantial demonstration in the literature[2-7] including direct blinded comparison to human planning with excellent performance. Like automated planning, autonomous vehicles

Department of Radiation Medicine and Applied Sciences, University of California, San Diego, La Jolla, CA.

Conflict of interest: None.

Address reprint requests to Kevin L. Moore, PhD, Department of Radiation Medicine and Applied Sciences, University of California, San Diego, La Jolla, CA 92093. E-mail: kevinmoore@ucsd.edu

https://doi.org/10.1016/j.semradonc.2019.02.003
1053-4296/

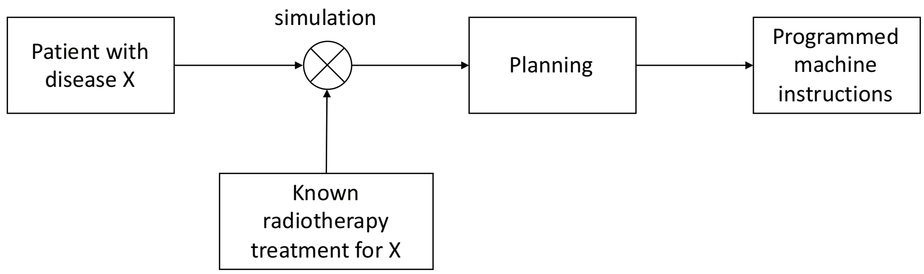

Figure 1 Simplified treatment planning process.

have some well-known barriers to adoption, including the difficulty in operating outside of standard conditions[8] as well as the upfront cost in implementing new systems.[9] Beyond the understandable barriers, the benefits of moving to fully autonomous vehicles are obvious:

- Mistakes made by human drivers could become a thing of the past.
- Human drivers could utilize their time in more productive ways.
- Redesigning transportation systems around self-driving cars could better manage finite resources (fuel, roads, air pollution).

Needless to say, each of these statements has a corresponding analogy in automated planning. There is, however, one point at which the analogy breaks down. No one expects self-driving cars to drastically decrease the time it takes to travel to a destination, while automating planning has the potential for extreme time reductions in the treatment design process.

In this article in Seminars, we will consider the current barriers to and the present opportunities for fully automated treatment planning. This discussion will utilize the most salient example at this moment in time: inverse-optimized external beam radiotherapy. While specific concepts such as the planning cost function will be framed in the context of intensity-modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT), the issue of patient-specific dose estimation and quality discrimination can be largely mapped over to other planning domains.

Why is Planning Difficult?

Before considering how automation can overcome the challenges posed by the planning process, a proper accounting of these challenges is essential. Ultimately, treatment planning is difficult because every patient presents a unique geometric arrangement of anatomic structures that must be treated with a correspondingly unique set of machine instructions. While there can be substantial interpatient similarity in some disease sites, there is more than enough anatomic variation between patients to confound most decisions regarding whether a candidate plan has achieved a requisite level of quality.[10]

The difficulty in planning is ultimately an issue of quality discrimination, that is how to assess good plans as "good" and bad plans as "bad." The definitions of good and bad plans are naturally patient-focused but also critically patient-specific: a good plan maximizes the therapeutic ratio *for that patient* and a bad plan fails to meet achievable clinical goals *for that patient*. There are myriad ways to err in the treatment planning process that would yield a suboptimal planning result:

- Errors in simulation
  - Incorrect representation of the patient.
  - Incorrect representation of radiation distribution in patients.
- Errors in contouring
  - Incorrect target delineation.
  - Incorrect normal tissue delineation.
- Errors in plan optimization
  - Planner fails to meet achievable plan quality for patient.
- Errors in delivery
  - Treatment delivery differs from representation in the treatment planning.

Simulation and delivery have generally "patient-independent" QA processes,[11-15] in that systems such as dose calculation verification and IMRT QA ensure the integrity of individual treatment plans by applying universal criteria to all patient plans. For example, while it must be assured that an individual patient's simulation scan utilized an established scanning protocol and their final modulated fields passed standard quality criteria,[14,16] these standards are unchanged between equivalent patients with the same disease and treatment. Contouring and plan optimization, however, are highly patient-specific and thereby far more difficult to standardize. The lack of patient-specific gold standards in anatomical delineation is well-known and has already been discussed in the preceding article on Autosegmentation in this issue. Thus for this Seminar, we are left with plan optimization as the primary patient-specific error to consider.

For this, we can advance a working definition of treatment plan quality:

> *Given a desired therapeutic dose of radiation to* ***a patient****, treatment plan quality is the degree to which a dose distribution maximizes tumor control and minimizes normal tissue injury* ***for a given technique****.*

This definition can be broken down into general features (underlined) and **specific features** (bolded). General features of plan quality are those components of a treatment that are common to all patients with the same disease. These general features are informed by the current state of knowledge in the field regarding response of target and normal tissue to the standard course of radiation for that disease. General features are subject to change as we learn more from evidence-based studies about the best way to treat our patients.

Specific features of plan quality can be broken down into 2 components: delivery technique and patient anatomy. The delivery technique may feel like a general feature, but it allows for different plan quality frontiers between, for example, 3D conformal and intensity-modulated radiotherapy, when treating the same disease. (If the technique is held fixed, eg, VMAT with a C-arm linear accelerator and 5 mm multileaf collimators, then this ceases to be a feature at all and is just taken to be the standard of care for this disease.) The most important specific feature that determines plan quality is thus the patient anatomy, which ultimately constrains a clinician's ability to achieve some or all of the desired general features of plan quality.[17-20] Unfortunately, the precise bounds for any given patient are nontrivial to determine quantitatively, meaning human planners must utilize trial-and-error to navigate toward the limiting tradeoff between the various competing parameters. The most common competing parameters are risk of normal tissue complication and tumor control probability, but can also manifest as balancing one normal tissue complication against another involved organ in the treatment field. Treatment planning is hard because the line between a good and bad plan must be uniquely negotiated for each patient, and most treatment planning processes do not sufficiently constrain the possible solution space.

Thus, the most unfortunate corollary to the fact that treatment planning is difficult is that not every treatment plan achieves a sufficient level of quality.[21] Stated another way, because it is difficult to objectively score individual treatment plans then human planners must utilize other means to assess quality. These subjective quality assessments can be very good, particularly when human pattern recognition is used to identify features of a high-quality dose distribution, for example the expected distance to the 50% isodose line. However, these subjective quality assessments can also fail, particularly when the quality space is multicriterial (as nearly all radiotherapy plans are). To focus on the *dose-volume histogram (DVH)* as the vehicle for quality assessment, a restatement of this problem comes by observing that good DVHs on unfavorable patients can be mistaken for bad DVHs because they score poorly against general (static) quality markers, while bad DVHs on favorable patients can be mistaken for good DVHs because they score adequately against static quality markers. The former problem is primarily one of confusion and frustration, while the latter can have very negative consequences for patients because they might receive far more dose to critical structures than necessary. One of the most salient examples of this occurred on the RTOG 0126 clinical trial in intermediate-risk prostate cancer, where the "best" treatment plans were observed in the cases where it was difficult to meet static protocol objectives while the "worst" treatment plans were found in instances where it was relatively easy to meet the static objectives and the planners stopped pushing the planning objectives to achieve further organ sparing.[21] If the last sentence appears paradoxical, the next section will provide an accounting to this nonintuitive result.

Patient-Specific Dose Estimations

One path out of the plan quality problem is to find a way to score treatment plans by obtaining some patient-specific "estimation" of a quantitative feature of the radiotherapy dose distribution. Possible features could include scalar dosimetric variables (eg, mean dose, V_x, D_y), DVHs, or the entire dose distribution itself. If sufficiently accurate, these patient-specific dose estimations allow a candidate treatment plan to be quantitatively compared to the expected values of some set of dose quality metrics. Assuming $i = 1\text{-}N$ quality metrics of interest, such a formulation would allow the following method to discriminate good and plan plans:

$$\begin{aligned} \forall_i \ \ if \ \left(D_i - D_{i,\ expected}\right) < \Delta_i \ \ &\rightarrow plan\ is\ good \\ else \ \ &\rightarrow plan\ is\ bad \end{aligned} \tag{1}$$

where $\forall_i$ means for all instances of i, D_i is the candidate plan's value for the i^{th} quality metric, $D_{i,expected}$ is the expected value for the i^{th} quality metric, and Δ_i is the error tolerance in the same. Positing such a framework is of course easier than obtaining patient-specific values of $D_{i,expected}$ and Δ_i; practically, this must occur by connecting anatomical features to dosimetric features. The following anatomical parameters have been shown to be predictive of dosimetric parameters:

- Target volume[18,22,23]
- Target shape[23]
- Relative location of target[17]
- Percentage of organ overlapping the *planning target volume (PTV)*[10]
- Distance-to-target/overlap volume histogram for organs-at-risk[17,18,20]
- Beam projections to targets and organs[22,23]
- Relative distance parameters between organs.[22]

Many of these parameters are not available to the human planner in standard treatment planning systems, and even when they are the connection to $D_{i,expected}$ and Δ_i must be established with some data-driven system. To this end, several strategies could be employed to establish how geometric features can be used to obtain accurate patient-specific dose estimations.

Heuristic Dose Estimation

In this method, patient-specific dose predictions are obtained by estimating dose-geometry relationships by use of some semiempirical model, for example by applying

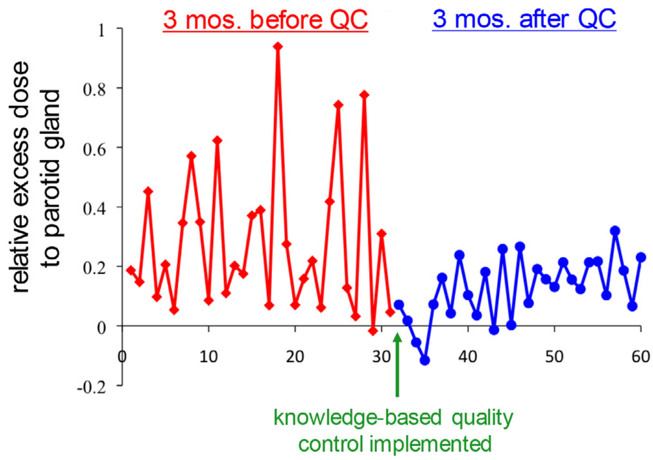

Figure 2 Clinical effect of patient-specific dose estimation on treatment plan quality. In Moore et al.,[10] a simple knowledge-based dose estimation model predicting mean parotid gland dose was used for plan quality control (QC) and had a dramatic effect on the excess dose ($[D_{manual} - D_{estimated}]/D_{estimated}$) given to parotid glands in real-world head-and-neck planning.

observed achievable dose gradient values outside target boundaries in a radiotherapy treatment. With such a heuristic method, dosimetric quantities of interest (eg, organ DVHs) can be obtained by estimating voxel doses outside of target(s) with a model of dose fall-off.[24] It should be noted that this method is data-driven in the sense that it utilizes observed beam behavior to tune the estimation model, but does not require previously-treated cases as training.

Library-based Dose Estimation

Library-based dose prediction is an intuitive method where by a cohort of previously-treated patients are available to compare to a new patient. To obtain a prediction for a new patient, a similar previous patient (or set of similar patients) could be matched to the new patient by utilizing some geometric similarity metric. Critical to this enterprise is the method of matching patients, which must utilize one or more of the dose-predicting geometric features above to connect new patients to previous observations. The utility of this method was demonstrated by Wu et al.,[20] whereby achievable DVH estimations were obtained by navigating a library of patients using the overlap-volume histogram to match prior observations to new cases. This method can be utilized to obtain dose estimations and/or automated planning objectives, though care must be taken to ensure that any previously suboptimal plans in the database will not have a negative effect on cases going forward.

Model-based Dose Estimation

This method is the engine behind what is commonly known as knowledge-based dose estimation, whereby patient specific dose predictions are obtained by training a mathematical model through statistical inference upon multiple prior patient observations.[10,17,18,22] The geometric variables can be similar, if not identical, to the matching variable utilized in library-based techniques. The difference comes in that, once trained, the dose estimation model makes no further use of the training database. The dose estimation model validation process, which mimics closely the automated planning validation process described later in this Seminar, can also obtain reasonable estimation error values Δ_i by checking the $D_{i,observed} - D_{i,expected}$ differences in a known high-quality validation set.

However they are obtained, patient-specific dose estimations in this way to discriminate high vs low quality plans can be fairly deemed “plan quality control” because the action applied universally to all patients in a multipatient planning context has the effect of yielding a more standardized, high-quality plans on average. This was demonstrated in Moore et al. whereby a simple mean dose prediction model allowed a previously-unavailable plan scoring method to perform quality discrimination in head-and-neck plans to dramatic effect (Fig. 2).[10]

Recent developments in model-based dose prediction have focused on utilizing ever larger training sets and ever more sophisticated methods[22,25] to make voxel-level estimations of expected dose. Such approaches exploit the rapidly-advancing field of machine learning to make ever more headway into the patient-specific prediction space, and as Deep Learning[26] methods become more commonplace the accuracy and specificity of these prediction systems are likely to increase substantially. Accurate voxel-level dose predictions can facilitate highly-nuanced plan quality discrimination,[23] and form the near ideal predicate to automated planning with patient-specific dose estimations.

Automated Planning With Patient-specific Dose Estimations

Intuitively, utilizing patient-specific dose estimations to check whether candidate plans are acceptable is a highly inefficient process, particularly if effort was involved in developing a plan that was ultimately found to require further manual optimization. Far superior would be the utilization of these estimations as guidance for plan optimization,[5,17] whereby no time is wasted in the service of generating categorically low-quality plans.

As current inverse optimization algorithms are all DVH-based, utilizing patient-specific DVH estimations to automatically generate appropriate patient-specific IMRT optimization objectives yields a systematic means to automated planning. DVH estimations, be they knowledge-based, transferred from matched prior patient(s), or heuristically-derived, are necessary but not sufficient conditions for this type of automated planning. Presently, the specification of *how* DVH predictions are required in the construction of optimization cost functions. Returning to the quality discrimination method of Eq. (1) and assuming the dose quality metrics are composed of DVH-derived quantities, one intuitive method for how to proceed could be posited by placing optimization objectives at each of the i=1-N DVH locations

$$D_{i,objective} \rightarrow (D_{i,expected} - \Delta_i) \quad (2)$$

This method would, on average, create a set of DVH points that are just out of reach of the achievable final dose distribution. This tracks with a standard plan optimization principle that objectives should be placed so that the DVH set points are beyond (but not too far beyond) the final plan value so that each contribution to the cost function is non-zero (but not too non-zero).

All that's missing to create a unified automated planning framework is the other element of an IMRT cost function, that is the respective weight of each DVH objective. Any given DVH estimation-based automated planning methodology could utilize the dose predictions and prioritization differently, as there are myriad options and any one treatment planning system's specific weight and/or priority algorithm need be accounted for. One intuitive approach for an objective weighting scheme would be to score of each of the N objectives according to their respective clinical priority and how the expected value compares to an accepted threshold for that metric (if one exists). Assuming a classically-formulated quadratic cost function

$$cost = \sum_{i=1}^{N} w_i \times \left[D_i - (D_{i,expected} - \Delta_i)\right]^2 \quad (3)$$

the weight w_i could be a function of several factors:

$$w_i \rightarrow \begin{cases} relative\ clinical\ importance\ of\ D_i \\ D_{i,expected} \\ D_{i,threshold} \end{cases} \quad (4)$$

The best schema for specifying objective placement and weighting will ultimately be determined by the characteristics of the *treatment planning system (TPS)* optimization framework and the validation testing discussed later in this Seminar.

Dynamic Inferred Parameter Optimization

Automated planning is possible without patient-specific dose estimations, as long as the optimization cost function is tuned in some way such that the final set of objectives and weights are ultimately patient-specific. One way to accomplish this could be termed *dynamic inferred parameter optimization*, whereby a cost function is continuously tuned depending on the value of each cost function element.[27] The following pseudo-algorithm would allow for an automated progression of a generalized cost function given by

$$cost = \sum_{i=1}^{N} w_i \times \left[D_i - D_{i,set}\right]^2 \quad (5)$$

where the initial values are given by

$$D_{i,set} \rightarrow D_{i,0},\ \ w_i \rightarrow w_{i,0} \quad (6)$$

At the end of each optimization iteration, the DVH set values and weights are adjusted according to some rubric

$$\begin{gathered} \forall_i\ if\ w_i \times (D_i - D_{i,set})^2 > \gamma_i^{upper} \rightarrow increase\ D_{i,set}\ \text{and/or}\ decrease\ w_i \\ elseif\ w_i \times (D_i - D_{i,set})^2 \gamma_i^{lower} \rightarrow decrease\ D_{i,set}\ \text{and/or}\ increase\ w_i \\ else \rightarrow do\ nothing \end{gathered} \quad (7)$$

The final values of $D_{i,set}$ might not arrive at the same values as $D_{i,expected}$ from the preceding framework, but they will be uniquely determined on a patient-by-patient basis. If $D_{i,expected}$ is available, this approach could be combined with a dose estimation-based framework to facilitate faster convergence.

Multiple Prospective Planning

One particularly elegant method for solving not only the patient-specific planning problem but also the problem of navigating the trade-offs inherent in the radiotherapy planning process is multicriterial optimization (MCO). MCO is a mature technique,[3,28-30] which can be briefly characterized as a method of generating a large space of prospective plans that explores the variance of the multiple clinical objectives, for example D_i. The advantage of this method is that it puts the tools of balancing the various criteria into the hands of the clinician, and the effect of altering one objective on the competing objectives is directly observable. The downside of this method to the task at hand is that it is not immediately automated. The tradeoff space is designed to be tuned by the clinician, which introduces manual decision-making back into the planning process, albeit in a much more constrained and quality controlled fashion than prior incarnations of unconstrained, user-specified DVH objective-based inverse planning. It is possible to select amongst the space of prospectively-generated plans the instance that best meets a finite list of criteria $D_{i,\ desired}$. Properly implemented, this strategy would preserve the option of navigating away from the automatically generated solution in nonstandard cases, while obviating the need for manual adjustment in most cases.

How to Test and Validate Automated Planning Routines

To assess how automated planning will perform against existing manual processes, the primary validation process is the application of the automated planning routine against a representative sample of previously treated patients. Whenever possible, these data should be separate from any knowledge-based training set, though this is sometimes not possible if training data is scarce. The rationale for a multipatient sample is obvious: to truly map out how the automated planning system compares to previous experience, it will need to demonstrate wide applicability to a range of geometric variability. As the set of applicable quality metrics can be quite large, summarizing them all in a succinct manner is a challenge. One technique is to simply lay out statistical summaries of $D_{i,manual} - D_{i,autoplan}$ in a table, including some measure of statistical significance where the null hypothesis is that automated planning is no different than manual planning for that parameter. Table 1 shows a simple case of this for VMAT prostate planning with a 53-patient validation set:

The statistically significant parameters observed in Table 1 form testable predictions for how the system should perform going forward. Interpreting one column of the table, the results of this automated planning validation test can be read as saying that, on average, automated plans in the future will be likely to reduce rectum V_{40Gy} if this technology supplants the manual process that created the validation set.

A related summarization technique of displaying all structures' DVH differences ($DVH_{manual}(D) - DVH_{autoplan}(D)$) is also useful to visualize the multipatient, multiparameter set. In Figure 3, the case of a newly-developed glioblastoma VMAT automated planning routine is shown.

The advantage of utilizing DVH difference visualization is that it can highlight outlier instances better than summary statistics. For example, in the case of the brainstem in the glioblastoma cohort most of the automated planning results matched prior experience with the exception of 2 plans that differed substantially in the critical 50-60 Gy region. These 2 apparent outliers would necessitate scrutiny of these individual cases to see why the gap manifested between the automated planning and the prior manual planning. For the example given, the root causes were sub-optimal previous planning (in the case where the autoplan was superior) and a total encompassing of brainstem with the PTV (in the case where the manual plan appeared superior). The former is an expected result, while the latter is indicating geometric conditions where the automated planning routine might not perform as well.

It is possible that when evaluating an automated planning that previously presupposed quality features need to be reasserted. The modulation factor for a VMAT and/or IMRT is one example, which is not a feature of the dose distribution but rather a condition of delivery quality. It has been

Table 1 Example Performance of an Automated Planning Routine in Prostate Cancer. In This Case, 53 Previously Treated Prostate VMAT Plans Were Replanned With a Knowledge-based Automated Planning Routine. A Set of Standard Dose Metrics Were Compared Across the Cohort, With the Difference $D_{manual} - D_{autoplan}$ quantifying the relative Percentage Difference Between the Manual Plans and The Automated Plans. Positive Values Indicate Improvement in the Automated Planning Group. Paired t-tests Quantified Statistical Significance in Each Metric, With n.s.: $P \geq 0.05$, *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$. Credit: Robert Kaderka, Ph.D.

Prostate (Rx = 81 Gy)	PTV $D_{100\%}$	PTV $D_{1\%}$	Bladder V_{40Gy}	Bladder V_{65Gy}	Bladder V_{75Gy}	Lt Femur Dmax	Rt Femur Dmax	Penile Bulb Dmean	Rectum V_{40Gy}	Rectum V_{65Gy}	Rectum V_{75Gy}
Dose metric difference [%]	0.0	0.2 ± 1.2	2.2 ± 2.6	0.3 ± 1.0	0.1 ± 0.7	−1.2 ± 6.0	−1.1 ± 8.1	6.6 ± 6.9	4.7 ± 4.9	1.4 ± 1.9	1.0 ± 1.2
Statistical significance	n.s.	n.s.	***	*	n.s.	n.s.	n.s.	***	***	***	***

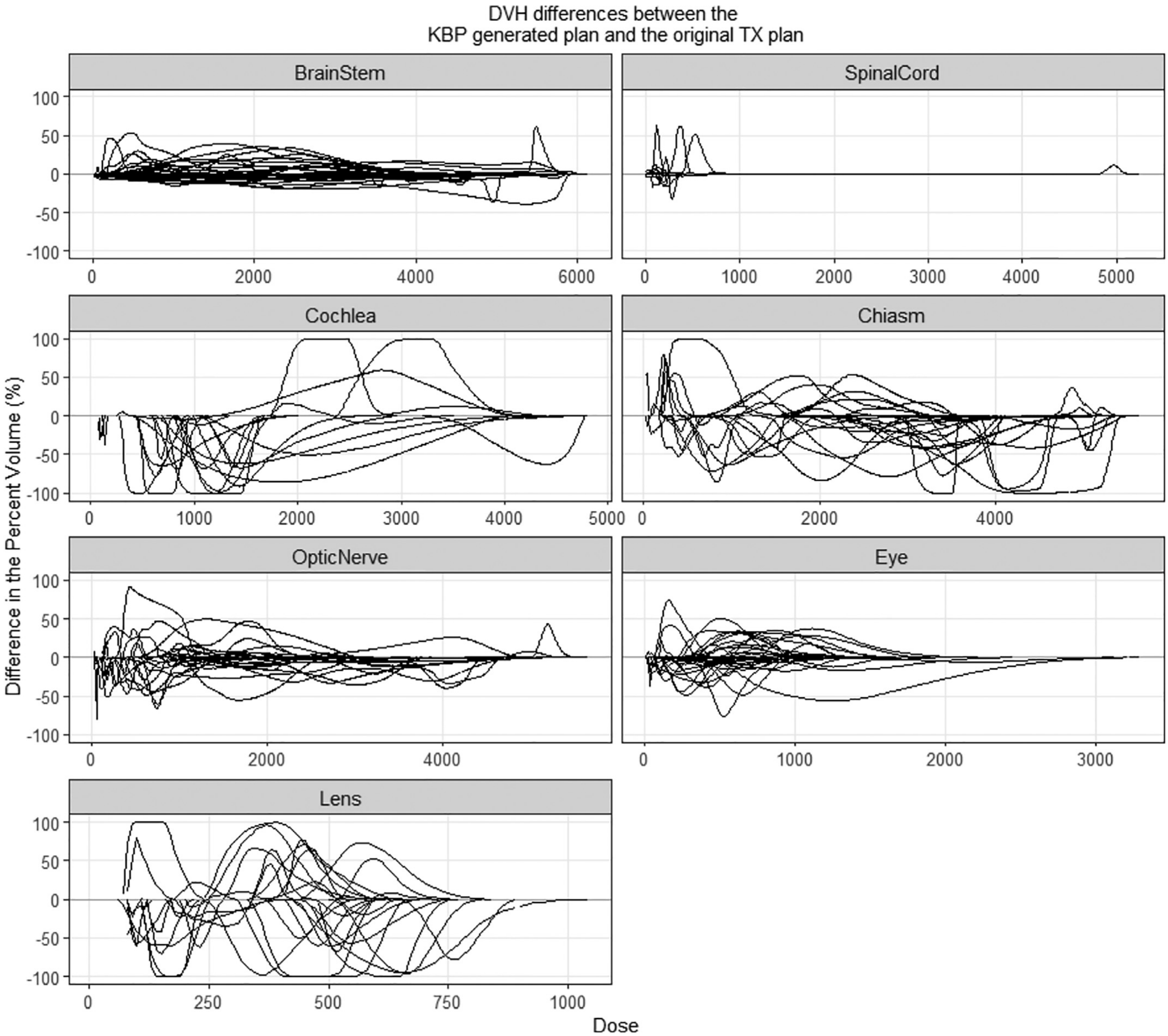

Figure 3 Multipatient, multiparameter evaluation of automated planning performance against prior clinical datasets for a newly-developed VMAT glioblastoma automated planning routine. The cumulative DVH difference trajectories $DVH_{manual}(D) - DVH_{autoplan}(D)$ for each OAR show how the automated planning routine. Credit: Xenia Ray, Ph.D.

observed that automated planning incarnations have systematically increased the modulation factor, which could be missed if this is not examined systematically. Dosimetric examples include placement of the maximum dose inside the target and the compactness of the 50% isodose line, which might not appear on a clinician's initial list of dosimetric quality features yet nevertheless could be noticeably different in an automated planning solution that was narrowly focused on improving a set of quality features that did not include these factors. This speaks to a general principle that all dosimetric quality features should be quantified in some way so that automated planning algorithms can factor this into their quality scoring methodology.

Beyond evaluation of each individual quality metric, an attractive complementary evaluation strategy is a blinded side-by-side comparison of automated and manual plans. The utility of blinding the plan identity is obvious, acting as the radiotherapy equivalent of a Turing test for treatment design.[31] If the output of the automated planning system is indistinguishable from the manually generated plans, then any concerns about adopting automated planning can be allayed by having verified that the system is noninferior, if not outright superior, to existing processes. As conducting a blinded retrospective review of many prior patients could be too much of a burden incorporating blinded review into an automated planning implementation strategy ensures both safe operation and workforce confidence in the system.[2]

One final consideration regarding clinical validation of automate planning is the general issue of attempting to infer the behavior of a system from a finite set of observations. Even if the system has generated clinically acceptable plans N times, one is never assured that the $(N+1)^{th}$ case will not have as-yet unobserved problems. The character of this type of problem is the subject of the next section, but one important conceptual element of the validation process is to constrain the likelihood of such automated planning errors. When this system is performing within the range of anatomical variation observed in the training and/or validation set, the output can be better assured than when the system is required to extrapolate to a geometric situation not yet observed (Fig. 4). Whenever new cases present themselves that are significant geometric outliers with respect to past experience, extra caution must be exercised in analyzing the output of the automated planning system.

Potential Drawbacks of Automated Planning

As all new technology must be taken skeptically, careful consideration of the limitations of automated planning is warranted. In this section we explore some of the potential problems facing automated planning as a definitive method for radiotherapy treatment design.

The "Enshrined Tradeoff" Problem

One of the selling points of a properly configured automated planning system is that it should consistently balance the same clinical tradeoffs across all patients. This can of course be a downside if new tradeoffs or new information need to be incorporated. Examples would include the addition of new avoidance structures (eg, regions identified by functional imaging) or the adjustment of previously understood quality markers (eg, updated *normal tissue complication probability (NTCP)* guidance parameters based on newly available outcomes data). These fundamentally alter the definition of a "good" plan, and the automated planning system would need to adjust accordingly. This can be easier said than done if the automated planning relies on a set of training plans that does not include this notion of good. In instances where automated planning must be retuned in this way, a significant amount of work could be in store for the practitioner in steering a new course for the automated planning system and validating it across multiple new patients.

The "Estimation Error" Problem

That an erroneous dose estimation could sometimes misguide the optimization is a critical concept to have in mind when using an automated planning system. Borrowing concepts from binary classifiers, these dose estimation errors can confound automated planning results in 2 distinct ways: false positives and false negatives. Using DVH estimation as the canonical example, a "Type I" estimation error could predict a DVH that is unachievable without a corresponding unacceptable offset to other clinical endpoints (most typically target coverage). This is depicted in Figure 5(A) whereby the blue dotted-line prediction is far lower than the achievable "best" DVH value shown as a red dashed-line. The resultant plan will likely fall somewhere in the middle, not achieving the impossible but aggressively over-sparing the organ and undermining the plan elsewhere. Strong secondary plan quality control that monitors all structures' quality parameters will likely detect that something has gone awry, though it will not necessarily be easy to discern the root cause of the confusion because, of course, no one has full knowledge of the "best" DVH that is depicted. How such an error would have to be corrected is the subject of the next subsection.

Far more insidious to the enterprise of automated planning are "Type II" errors whereby the dose estimation yields a prediction that is too high, that is the achievable "best" DVH lies below the estimation as in Figure 5(B). This is a particularly troublesome possibility because putting in an easily met DVH objective on one structure will incur no negative consequences to other structures, meaning the detectability of this error is substantially less than Type I errors. As described in the previous subsection, understanding the range of error in the validation set becomes critical for appreciating the bounds of this possibility. It is also the case that this type of error is likely only when one or more anatomical parameters are well outside the range of the training set distribution as in Figure 4, meaning there could be signals to the user that this is more or less probable.

To combat Type I and II estimation errors, it would be possible to implement a redundant dose estimation system, akin to the well-known secondary dose calculation verifies monitor units in clinical scenarios. Ideally this system should utilize an entirely different estimation algorithm to provide a truly independent check against the primary system.

The "Correction" Problem

If one is able to discern any Type I/II errors in the automated plan, or if a different set of clinical tradeoffs need be made for a singular case, the task of adjusting a treatment plan must be contended with. Depending on the implementation of automated planning, this task might necessitate returning to the very system of manual adjustment of optimization parameters that automated planning sought to replace. To truly realize both the advantages of automated planning with the freedom to fine tune when warranted, new tools of adjustment should be deployed. Multicriterial optimization is an excellent example of how to navigate the tradeoff space, as is the nascent ability to directly adjust the isodose distribution.[32] As a general principle, these new adjustment tools should be designed with the attending physician in mind so that the clinicians with the most direct knowledge of what tradeoffs need be made have the ability to do so.

The "Skill Atrophy" Problem

Closely connected to the Correction Problem, a long-term reliance on automated treatment planning will surely have the effect of depleting the skills of a workforce to plan "the old fashioned" way. The necessity of maintaining manual planning experience of existing personnel and educating

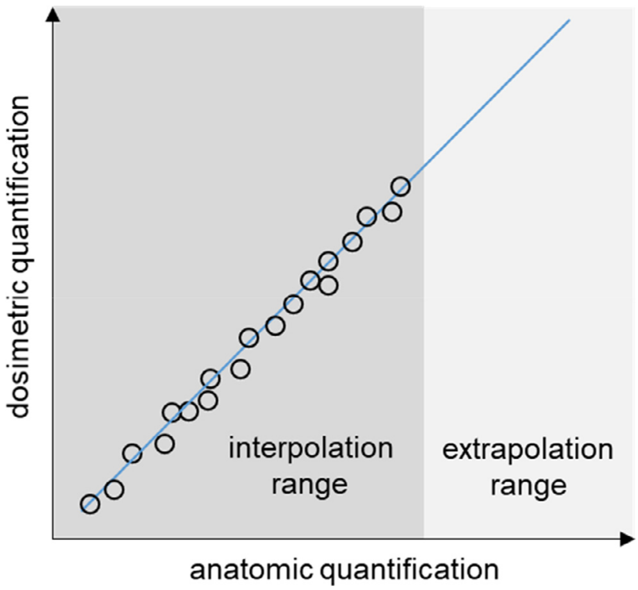

Figure 4 Diagrammatic representation of dose estimation and automated planning representing the observed range of prior experience. Future performance in the anatomic range outside of prior observations can be extrapolated, but the accuracy of the dose estimation and/or automated planning system in such an instance remains untested.

new trainees in the era of automated planning remains to be determined. There is little need in a modern radiotherapy department for staying current on photon block cutting procedures, however the ability to override and adjust automated image registration remains an essential skill.

Quality Assurance and Ongoing Maintenance of Automated Planning Routines

As automated planning routines subsume the more mundane tasks of designing treatment plans, they are ultimately but one more clinical system that must be tested and maintained as part of a quality management program. Analogous to beam models, planning modules must be verified through software version upgrades and adjustments to ancillary software components (eg, dose calculation models, optimization algorithm changes). Similarly, an efficient means by which to verify performance of an automated planning could include a benchmark dataset, potentially even the same validation set used at the time of commissioning.

As some knowledge-based automated planning routines benefit from broadening the training set, routine incorporation of new data into the dose estimation models is warranted, particularly if the new cases are either achieving better results than previous training cases or represent extreme geometric observations outside the previously observed range of anatomic variation. Obviously as new delivery techniques, for example 4π radiotherapy, displace previous methods then the automated planning systems will have to adapt their core dose estimation and validation performance. Each new instance of an automated planning routine should be benchmarked as if it were a new system. When replacing one automated planning routine for another, the validation set used to the test the new system could, and perhaps should, remain fixed to allow easy comparison between the two automated systems.

The Future of Planning: Does It Have One?

Clinical processes evolve. There are numerous human-driven tasks previously essential to the radiotherapy process that is no longer required. Photon block cutting, physical compensator construction, manual parameter entry on the linear accelerator; these tasks used to occupy a significant amount of manual clinician effort yet have been replaced by automation. It is very much an open question whether or not human planners will be required for routine treatment design in the near future (Fig. 6).

But the question of whether planning has a future is a deeper one even than the narrow consideration of whether human beings will be adjusting optimization parameters in the next 10 years. Central to the notion of treatment planning is the assumption that something must happen between contouring and an assessment of the treatment quality.

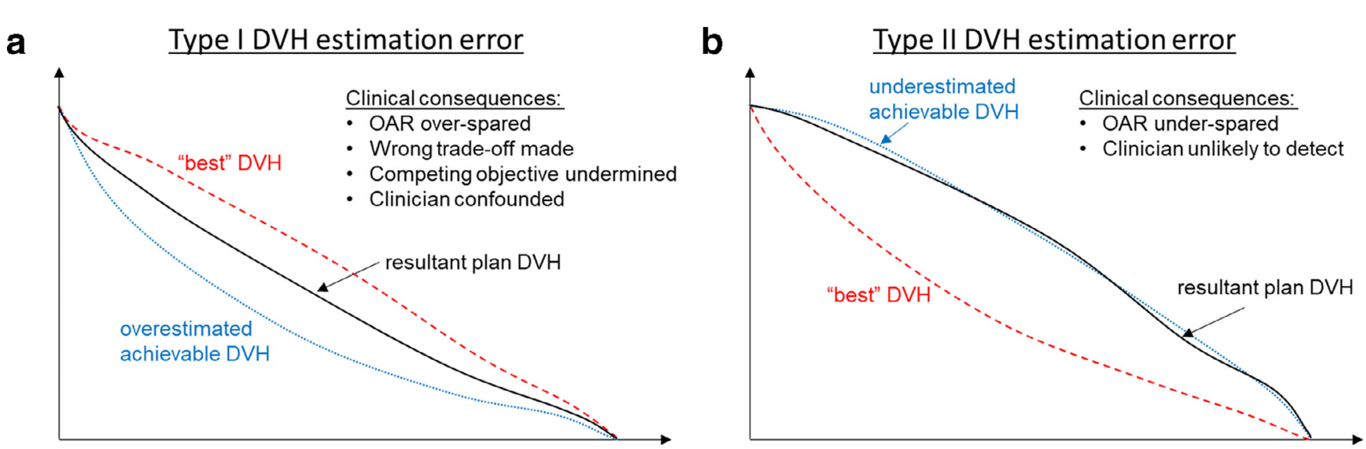

Figure 5 DVH estimation errors and their clinical consequences.

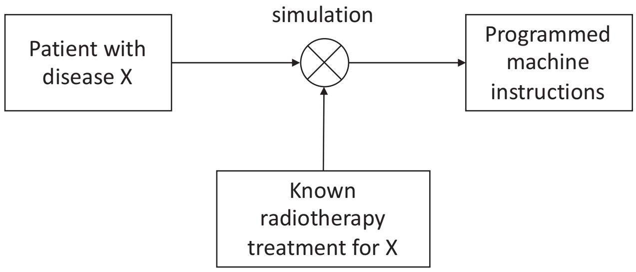

Figure 6 Simplified treatment design process of the future, whereby the planning task has to be subsumed into the simulation step.

Why must this be the case? If a highly-accurate 3D dose distribution estimate is *currently* available in less than a second[23] how long will it be before a deliverable machine sequence is available on the same timescale? At some point in the future, a physician will complete her last contour and instantaneously a deliverable dose distribution will appear on-screen. And at some point not long thereafter, the thought of there being any appreciable delay between target delineation and dose review will be unthinkable. "Show dose" will become one more button in the Treatment Planning Syst... well, a new name for that bit of antiquated software will likely be required.

References

1. Teoh ER, Kidd DG: Rage against the machine? Google's self-driving cars versus human drivers. J Safety Res 63:57-60, 2017
2. Ziemer B, Shiraishi S, Hattangadi-Gluth J, et al: Fully-automated, comprehensive knowledge-based planning for stereotactic radiosurgery—pre-clinical validation through blinded physician review. Pract Radiat Oncol 7:e569-e578, 2017
3. Li N, Carmona R, Sirak I, et al: Validation of a Knowledge Based Automated Planning System in Cervical Cancer As a Clinical Trial Quality System. San Antonio, TX: Annual Meeting of the American Society for Radiation Oncology; 2015
4. Voet PW, Dirkx ML, Breedveld S, et al: Toward fully automated multicriterial plan generation: A prospective clinical study. Int J Radiat Oncol Biol Phys; 2012
5. Quan EM, Chang JY, Liao Z, et al: Automated volumetric modulated arc therapy treatment planning for stage III lung cancer: How does it compare with intensity-modulated radio therapy? Int J Radiat Oncol Biol Phys 84:e69-e76, 2012
6. Wu B, McNutt T, Zahurak M, et al: Fully automated simultaneous integrated boosted—intensity modulated radiation therapy treatment planning is feasible for head-and-neck cancer: A prospective clinical study. Int J Radiat Oncol Biol Phys; 2012
7. Purdie TG, Dinniwell RE, Letourneau D, et al: Automated planning of tangential breast intensity-modulated radiotherapy using heuristic optimization. Int J Radiat Oncol Biol Phys 81:575-583, 2011
8. Saikin AM, Buznikov SE, Shabanov NS, et al: The mathematical model of dynamic stabilization system for autonomous car. IOP Confer Ser: Mater Sci Engineer 315:012023
9. Fagnant DJ, Kockelman K: Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transport Res Part A 77:167-181, 2015
10. Moore KL, Brame RS, Low DA, et al: Experience-based quality control of clinical intensity-modulated radiotherapy planning. Int J Radiat Oncol Biol Phys 81:545-551, 2011
11. Kutcher GJ, Coia L, Gillin M, et al: Comprehensive QA for radiation oncology: Report of AAPM Radiation Therapy Committee Task Group 40. Med Phys 21:581-618, 1994
12. Klein EE, Hanley J, Bayouth J, et al: Task Group 142 report: Quality assurance of medical acceleratorsa). Med Phys 36:4197-4212, 2009
13. Fraass B, Doppke K, Hunt M, et al: American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: Quality assurance for clinical radiotherapy treatment planning. Med Phys 25:1773-1829, 1998
14. Miften M, Olch A, Mihailidis D, et al: Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys 45:e53-e83, 2018
15. Mutic S, Palta JR, Butker EK, et al: Quality assurance for computed-tomography simulators and the computed-tomography-simulation process: Report of the AAPM Radiation Therapy Committee Task Group No. 66. Med Phys 30:2762-2792, 2003
16. Low DA, Harms WB, Mutic S, et al: A technique for the quantitative evaluation of dose distributions. Med Phys 25:656-661, 1998
17. Appenzoller LM, Michalski JM, Thorstad WL, et al: Predicting dose-volume histograms for organs-at-risk in IMRT planning. Med Phys 39:7446, 2012
18. Zhu X, Ge Y, Li T, Thongphiew D, et al: A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. Med Phys 38:719-726, 2011
19. Wu B, Ricchetti F, Sanguineti G, et al: Data-driven approach to generating achievable dose—volume histogram objectives in intensity-modulated radiotherapy planning. Int J Radiat Oncol Biol Phys 79:1241-1247, 2011
20. Wu B, Ricchetti F, Sanguineti G, et al: Patient geometry-driven information retrieval for IMRT treatment plan quality control. Med Phys 36:5497-5505, 2009
21. Moore KL, Schmidt R, Moiseenko V, et al: Quantifying unnecessary normal tissue complication risks due to suboptimal planning: A secondary study of RTOG 0126. Int J Radiat Oncol Biol Phys 92:228-235, 2015
22. Shiraishi S, Moore KL: Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. Med Phys 43:378-387, 2016
23. Yuan L, Ge Y, Lee WR, et al: Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. Med Phys 39:6868-6878, 2012
24. Fried DV, Chera BS, Das SK: Assessment of PlanIQ Feasibility DVH for head and neck treatment planning. J Appl Clin Med Phys 18:245-250, 2017
25. C. McIntosh, Thomas GP: Voxel-based dose prediction with multi-patient atlas selection for automated radiotherapy treatment planning. Phys Med Biol 62:415, 2017
26. LeCun Y, Bengio Y, Hinton G: Deep learning. Nature 521:436, 2015
27. Tol JP, Dahele M, Peltola J, et al: Automatic interactive optimization for volumetric modulated arc therapy planning. Radiat Oncol 10:75, 2015
28. Jee KW, McShan DL, Fraass BA: Lexicographic ordering: Intuitive multicriteria optimization for IMRT. Phys Med Biol 52:1845-1861, 2007
29. Thieke C, Kufer KH, Monz M, et al: A new concept for interactive radiotherapy planning with multicriteria optimization: First clinical evaluation. Radiother Oncol 85:292-298, 2007
30. Craft D, McQuaid D, Wala J, et al: Multicriteria VMAT optimization. Med Phys 39
31. Turing Alan: Computing Machinery and Intelligence (PDF). Mind, LIX 236:433-460, 1950
32. Karl Otto: Real-time interactive treatment planning. Institute of Physics and Engineering in Medicine Physics in Medicine & Biology 59:17, 2014