Evolving treatment plan quality criteria from institution-specific experience

D. Ruan, W. Shao, J. DeMarco, S. Tenn, C. King, D. Low, P. Kupelian, and M. Steinberg[a)]
Department of Radiation Oncology, University of California, Los Angeles, California 90095

(Received 26 January 2012; revised 18 March 2012; accepted for publication 31 March 2012; published 23 April 2012)

Purpose: The dosimetric aspects of radiation therapy treatment plan quality are usually evaluated and reported with dose volume histogram (DVH) endpoints. For clinical practicality, a small number of representative quantities derived from the DVH are often used as dose endpoints to summarize the plan quality. National guidelines on reference values for such quantities for some standard treatment approaches are often used as acceptance criteria to trigger treatment plan review. On the other hand, treatment prescription and planning approaches specific to each institution warrants the need to report plan quality in terms of practice consistency and with respect to institution-specific experience. The purpose of this study is to investigate and develop a systematic approach to record and characterize the institution-specific plan experience and use such information to guide the design of plan quality criteria. In the clinical setting, this approach will assist in (1) improving overall plan quality and consistency and (2) detecting abnormal plan behavior for retrospective analysis.

Methods: The authors propose a self-evolving methodology and have developed an in-house prototype software suite that (1) extracts the dose endpoints from a treatment plan and evaluates them against both national standard and institution-specific criteria and (2) evolves the statistics for the dose endpoints and updates institution-specific criteria.

Results: The validity of the proposed methodology was demonstrated with a database of prostate stereotactic body radiotherapy cases. As more data sets are accumulated, the evolving institution-specific criteria can serve as a reliable and stable consistency measure for plan quality and reveals the potential use of the "tighter" criteria than national standards or projected criteria, leading to practice that may push to shrink the gap between plans deemed acceptable and the underlying *unknown* optimality.

Conclusions: The authors have developed a rationale to improve plan quality and consistency, by evolving the plan quality criteria from institution-specific experience, complementary to national standards. The validity of the proposed method was demonstrated with a prototype system on prostate stereotactic body radiotherapy (SBRT) cases. The current study uses direct and indirect DVH endpoints for plan quality evaluation, but the infrastructure proposed here applies to general outcome data as well. The authors expect forward evaluation together with intelligent update based on evidence-based learning, which will evolve the clinical practice for improved efficiency, consistency, and ultimately better treatment outcome. © *2012 American Association of Physicists in Medicine*. [http://dx.doi.org/10.1118/1.4704497]

Key words: plan quality, self-evolving dose criterion

I. INTRODUCTION AND PURPOSE

Quantitative evaluation and reporting is important for consistent generation of high-quality plans. Conventionally, a planner would visually examine the dose volume histogram (DVH) and judge against (1) standard planning coverage and organ at risk avoidance metrics; (2) dose criteria specified by the attending physician; and (3) the planner's personal experience as to the optimal DVH curve shape one may achieve for the specific pathology, patient geometry, and treatment modality. The criteria for (1) are often obtained from national standards or well-established reference literature.[1–13] However, standard plan quality criteria have their limitations. First, institution-specific implementation, such as difference in structure and/or margin definition, and potentially different prescription schemes with respect to volume, curve, or points may lead to different DVH trade-offs. Furthermore, the national standards could be easily achievable, and institutions equipped with advanced treatment machines and treatment planning systems may exceed such published standards—in such case, it is desirable for institutions to "shoot higher" rather than just satisfying the general criteria. Finally, published standards may be absent for relatively new or unique protocols. As a limitation regarding (2) and (3), the dose criteria set by the individual physicians and planners often rely on potentially biased personal experience and knowledge, susceptible to subjectivity and is desirable to be supported by quantitative evidence.

To overcome the aforementioned limitations of relying solely on published standards and personal experience, we propose to develop an institution-specific plan quality summary methodology. In addition to the conventional reporting

modules, plan quality statistics are collected based on existing plans of comparable pathology/treatment characteristics. Such statistics not only provide an institution-specific criteria table but also provide knowledge of the relative standing of a newly completed plan among similar cases. Furthermore, upon plan approval, the new plan is used to further update the institution-specific statistics. This self-evolving paradigm provides a valuable platform to examine plan consistency and to detect outlier plan quality values that may facilitates identification of either potential issues or further improvement on suboptimal plans.

II. METHODS AND MATERIAL

The key of this project is the dynamic generation of self-evolving plan quality criteria based on institution-specific experiences. Such metrics are statistical in nature and are meant to complement either the national standards or static criteria set *a priori* by the physicians. Figure 1 presents the general workflow for this process.

The plan quality report development consists of two major modules: "forward evaluation" and "criteria evolution." The forward evaluation module is most active when the goal is to obtain knowledge on the absolute and relative performance of a potential plan, e.g., at the stage when dosimetrists/physicists have generated an initial plan but have not sent it for physician approval. More specifically, the forward evaluation module consists of the following: (1) planner either informs the task management system or sends the plan information to a *distiller* where plan quality values of interest are extracted; (2) these values are simultaneously fed to two comparison processors, one against the standard static criteria (such as from RTOG if applicable) and the other against the institution-specific criteria from statistics of accumulated experience data. The criteria evolution module is evoked once a plan is approved and is considered as future reference value. At this point, the distilled plan quantity is incorporated to update the institution-specific statistics and subsequently the corresponding criteria. An in-house software suite is described below.

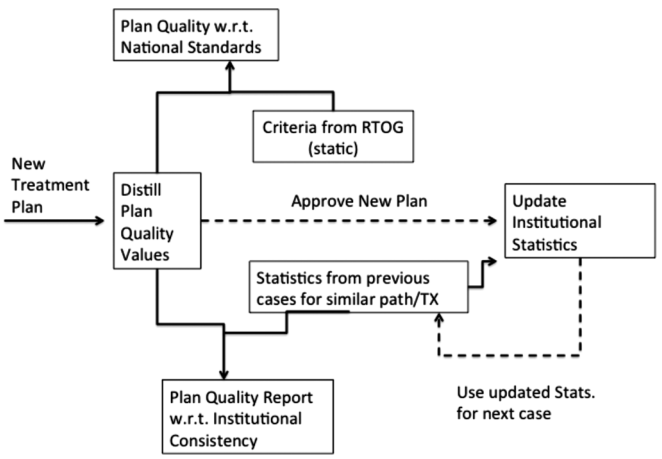

FIG. 1. Workflow for plan quality report, with respect to static (inclusive of RTOG-specified) criteria and self-evolving institutional statistics.

II.A. Action upon a new plan generation

We have explored a few options for the input component: it can be either actively pushed by the planner into the analysis system or pulled by the distiller. When plan quality is defined with direct DVH quantities, it is feasible to ask the planner to manually push the data field or export DVH curves for further analysis—this was the operation we adopted at the initial stage. At UCLA, the conventional clinical practice was to populate the fields of an excel sheet with values from the DVH curve evaluations directly from the treatment planning systems. As the project progresses, we have decided to alleviate the planner of the data pushing burden and instead operate on the DICOM-RT objects exported from various treatment planning system. To this end, we require the planners to routinely perform DICOM export to designated directories on a data server. We have developed a daemon program that runs in the background of the server and continuously monitors the DICOM destinations to detect addition and update of this directory. This functionality is currently implemented in JAVA, and we are currently developing shell daemon based on interruption signals.

II.B. Plan quality distiller

Once an DICOM-RT object is detected to have been created or modified in the file system, an in-house developed data processing interface, programmed in a combination of JAVA and C++, is used to detect the source treatment planning system for the DICOM-RT object and subsequently extract the image/structure/dose triplets by evoking treatment planning system specific interpretation code. The structure names are then mapped to a standard nomenclature set, utilizing a predefined dictionary and fuzzy logic matching with regular expressions. To support general purpose parsing for nonstandard plans with structure naming inconsistent with dictionary entries, an interactive user interface is also developed. Plan quality indices and conformality quantification, such as the ratio between high dose volume (e.g., above 50% prescription dose) and planning target volume, are distilled or calculated.

II.C. Representation and visualization of plan quality statistics

In an effort to have an informative yet compact representation for the statistics from experience data, we propose to present the median, and the quartiles $q_1 = 25\%$ and $q_3 = 75\%$. A "reasonable" range was defined by specifying the lower-bound l and upper-bound u to be $l \triangleq^{\Delta} q_1 - 1.5(q_3 - q_1)$ and $u \triangleq^{\Delta} q_3 + 1.5(q_3 - q_1)$, respectively, corresponding to approximately $\pm 2.7\sigma$ and 99.3% coverage if the data distribution was Gaussian. Values outside this range are considered as *outliers*, and individually visualized, as shown in Fig. 2.

II.D. Institution-specific criteria from plan quality statistics

Two reporting options are proposed to utilize the above statistics for plan quality evaluation. The first option thresholds the plan quality value with the corresponding percentile

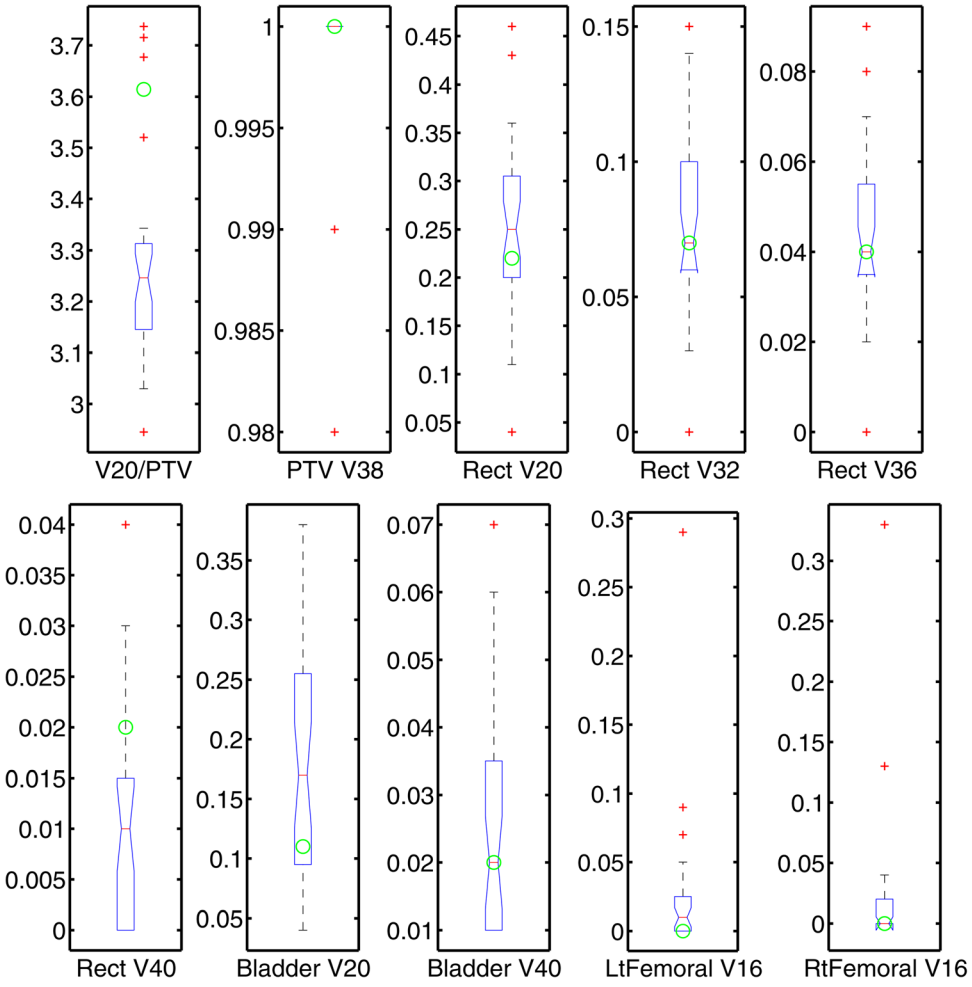

FIG. 2. Graphical plan quality report of a new plan evaluated using statistics from previous cases of similar pathology and treatment modality. Quantiles of 25% and 75% (ends of hour glass box), range limit values (whiskers), and outliers outside lower/upper bounds (crosses) are generated from all previously collected cases, while the plan quality evaluation for the newly planned case is presented with circles (°) for visual comparison.

statistics to trigger further plan review. More specifically, we use 25% percentile values for quantities where high values are desirable, such as planning target volume (PTV) coverage and use 75% percentile values for quantities desired to be low, such as organ at risk (OAR) dose volume. The rationale is to have a quick threshold criterion to identify plans with *unacceptable* quality, rather than optimal quality. The second option reports directly where the new plan stands relative to the existing statistics, where subsequent judgment is to be made by the reviewer.

II.E. User interaction

Since the generated statistics reflect the accumulated planning experience, it is desirable to enable users to interactively identify plans with abnormal performance. Since such investigation is usually directed toward outliers, our graphical interface displays the patient ID (MRN) when the cursor is placed on top of any outlier points.

III. RESULTS

The utility of the proposed system is demonstrated on prostate cases treated with stereotactic body radiotherapy (SBRT). The superior ability, compared to other radiotherapy techniques, of prostate SBRT has been demonstrated in its safety and its superior ability to spare rectum and bladder compared to the other radiotherapy techniques.[14,15] Our prostate SBRT protocol is a hypofractionated scheme with 40 Gy delivered over five fractions.

Table I lists the set of static DVH criteria used in our clinical protocol for prostate SBRT. Table II reports the DVH criteria obtained from the institutional experience based on 32 SBRT cases and evolving it to include a new approved case.

Figure 2 illustrates the collective statistics from 32 existing plans and the relative placement of the quality value for a new plan. The color coded box plot shows the statistics based on previous cases, with quantiles, extreme points, and outliers specified for each dose criterion. The quality of the new plan is presented on the same axis as the previous statistics to provide an intuitive visualization of its relative standing to the previous cases.

Comparing Tables I and II, we see that the quantile-based institution-specific review criteria (25% quantile value for V95 and 75% quantile value for the other endpoints) is uniformly more strict than the static criterion in Table I. Furthermore,

TABLE I. Static clinical criteria for prostate SBRT with prescription dose 40 Gy.[a]

Clinical variable	Clinical goal
PTV	
V38	≥95.0%
V20/VPTV	≤4.0
Rectum	
V20	≤50.0%
V32	≤20.0%
V36	≤10.0%
V40	≤5.0%
Bladder	
V20	≤40%
V40	≤10%
Left femoral	
V16	≤5%
Right femoral head	
V16	≤5%

[a]V# represents the relative volume (%) receiving more than #Gy of dose.

Fig. 2 shows the new plan is consistent with respect to the existing statistics, demonstrating its feasibility and the likelihood to be an acceptable plan. Upon acceptance of this new plan, its plan quality values are used to update the statistics, as reported in Fig. 3 and the right-hand-side columns in Table II. Note that this updated statistics will be used to determine the review criterion for the next plan. Figure 3 also illustrates the ability to allow the user to toggle the case ID (anonymized to protect personal information) by interacting with the graphical interface for retrospective examination. Comparing preupdate and postupdate statistics leads to the observation that incorporating a reasonably good plan that is consistent with previous cases hardly perturbs the statistics, as one would expect from theoretical analysis—as more consistent data points accumulate, the empirical statistics would converge to a set of stable values reflecting the underlying statistics.

IV. DISCUSSION

The benefit of establishing institution-specific criteria based on learning from past experience is several-fold. First, the collected performance is intrinsically consistent with the routine practice within the institution—in terms of structure contouring, margin design, prescription standards, normalization scheme, etc. The results from collecting and analyzing institution data alleviate the potential variation caused by such differences. In addition to minimizing the effect due to interinstitutional practice differences, the experience data also account for the characteristics of the specific treatment planning system, templates used for planning, techniques, and choice of algorithms specific to an institution. Second, as more quality data are accumulated, the law of large numbers becomes more valid, and the empirical distribution of the data one collects approaches the underlying statistical behavior of the plan quality variables asymptotically. Such knowledge is critical for the clinic to decide acceptance boundaries, as well as to obtain insight regarding the gap to optimality for a specific plan. Usually, failing to receive the attention it deserves, quantifying the level of suboptimality, is important in treatment planning strategy, in knowing when to call a plan "good enough" and cannot be further improved without further tradeoff among the clinically significant endpoints. Good acceptance and optimality criteria help the planners to design workflow protocol and strategies to balance plan quality and efficiency. Finally, accumulation and analysis of quality data within a clinic provides high-confidence (low variation) evidence for practice improvement and supports evidence-based directives on various levels.

The current study uses direct and indirect DVH endpoints for plan quality evaluation. We are working on extending current infrastructure to digest follow up results as they become available. Such hybrid information would work synergistically with the dose endpoints to provide insight into identifying variables of high prognostic power and the criteria associated with these variables. Forward evaluation together

TABLE II. Plan quality statistics pre- and postincorporation of the new plan into the cohort of 32 existing plans.

	Stat. based on existing plans			Stat. upon new plan incorporation		
Clinical variable	25% quantile	Median	75% quantile	25% quantile	Median	75% quantile
PTV						
V38	100%	100%	100%	100%	100%	100%
V20/VPTV	3.15	3.25	3.3	3.15	3.25	3.3
Rectum						
V520	20%	25%	31%	20%	25%	30%
V32	6%	7%	10%	6%	7%	10%
V36	3.5%	4%	6%	3.8%	4%	5%
V40	0	1%	1.5%	0	1%	1.5%
Bladder						
V20	9.5%	17%	25.5%	9.8%	17%	25.5%
V40	1%	2%	3.5%	1%	2%	3.3%
Left femoral						
V16	0	1%	2.5%	0	1%	2.3%
Right femoral head						
V16	0	0	2%	0	0	2%

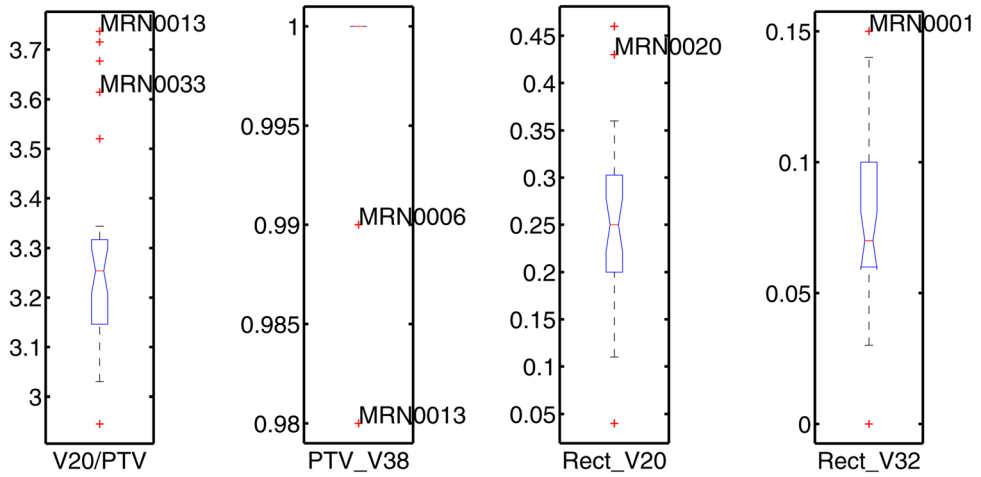

FIG. 3. Examples of graphical plan quality report for updated institutional plan quality statistics and interactively identified outliers with potential abnormality.

with intelligent update based on evidence-based learning will evolve the clinical practice for improved efficiency, consistency, and ultimately better treatment outcome.

ACKNOWLEDGMENTS

Dr. N. Agazaryan established the prostate SBRT physics program at UCLA. The authors also thank him for insightful inputs and discussions.

[a]Electronic mail: druan@mednet.ucla.edu

[1]B. Emami, J. Lyman, A. Brown, L. Coia, M. Goitein, J. E. Munzenrider, B. Shank, L. J. Solin, and M. Wesson, "Tolerance of normal tissue to therapeutic irradiation," Int. J. Radiat. Oncol., Biol., Phys. **21**, 109–122 (1991).

[2]L. S. Constine, M. T. Milano, D. Friedman, M. Morris, J. P. Williams, P. Rubin, and P. Okunie, "Late effects of cancer treatment on normal tissues," *Principles and Practice of Radiation Oncology* (Lippincott Williams & Wilkins, Philadelphia, 2008).

[3]E. Marks, R. J. Baglan, S. C. Prassad, and W. F. Blank, "Cerebral radionecrosis: Incidence and risk in relation to dose, time, fractionation and volume," Int. J. Radiat. Oncol., Biol., Phys. **7**, 243–252 (1981).

[4]S. A. Leibel, G. E. S. I. P. Gutin, S. Leibel, and G. Sheline, *Tolerance of the Brain and Spinal Cord to Conventional Irradiation* (Raven, New York, 1991).

[5]Radiation therapy oncology group available at http://www.rtog.org.

[6]M. L. Moster and R. Foroozan, "Complications of cancer therapy," *Walsh and Hoyt's Clinical Neuro-Ophthalmology* (Lippincott, Williams & Wilkins, Philadelphia, 2005), pp. 1759–1822.

[7]K. B. Gordon, D. H. Char, and R. H. Sagerman, "Late effects of radiation on the eye and ocular adnexa," Int. J. Radiat. Oncol., Biol., Phys. **31**, 1123–1139 (1995).

[8]C. A. Servodidio and D. H. Abramson, "Acute and long-term effects of radiation therapy to the eye in children," Cancer Nurs. **16**, 371–381 (1993).

[9]C. C. Pan, A. Eisbruch, J. S. Lee, R. M. Snorrason, R. K. T. Haken, and P. R. Kileny, "Prospective study of inner ear radiation dose and hearing loss in head-and-neck cancer patients," Int. J. Radiat. Oncol., Biol., Phys. **61**, 1393–1402 (2005).

[10]K. A. Leber, J. Bergloff, and G. Pendl, "Dose-response tolerance of the visual pathways and cranial nerves of the cavernous sinus to stereotactic radiosurgery," J. Neurosurg. **88**, 43–50 (1998).

[11]M. V. Graham, J. A. Purdy, B. Emami, W. Harms, W. Bosch, M. A. Lockett, and C. A. Perez, "Clinical dose-volume histogram analysis for pneumonitis after 3d treatment for non-small cell lung cancer," Int. J. Radiat. Oncol., Biol., Phys. **45**, 323–329 (1999).

[12]F. M. Kong, J. A. Hayman, K. A. Grith, G. P. Kalemkerian, D. Arenberg, S. Lyons, A. Turrisi, A. Lichter, B. Fraass, A. Eisbruch, T. S. Lawrence, and R. K. T. Haken, "Final toxicity results of a radiation-dose escalation study in patients with non-small-cell lung cancer: predictors for radiation pneumonitis and fibrosis," Int. J. Radiat. Oncol., Biol., Phys. **65**, 1075–1086 (2006).

[13]P. A. Kupelian, C. A. Reddy, T. P. Carlson, and T. R. Willoughby, "Dose/volume relationship of late rectal bleeding after external beam radiotherapy for localized prostate cancer: Absolute or relative rectal volume?," Cancer J. **8**, 62–66 (2002).

[14]C. Kingand, J. Brooks, H. Gill, T. Pawlicki, C. Cotrutz, and J. Presti, "Stereotactic body radiotherapy for localized prostate cancer: Interim results of a prospective phase ii clinical trial," Int. J. Radiat. Oncol., Biol., Phys **73**, 1043–1048 (2009).

[15]C. King, "Stereotactic body radiotherapy for prostate cancer: Current results of a phase II trial," *[IMRT IGRT SBRT—Advances in the Treatment Planning and Delivery of Radiotherapy]* Front. Radiat. Ther. Oncol. **43**, 428–437 (2011).