Published in final edited form as:

Phys Med Biol. ; 64(14): 145020. doi:10.1088/1361-6560/ab2dc5.

Task-based image quality assessment in radiation therapy: initial characterization and demonstration with computer-simulation study

Steven R Dolly¹, Yang Lou^{2,5}, Mark A Anastasio^{3,6}, Hua Li^{3,4,6}

¹SSM Health Cancer Care, St. Louis, MO, United States of America

²Facebook Reality Lab, Redmond, WA, United States of America

³Department of BioEngineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States of America

⁴Carle Cancer Center, Carle Foundation Hospital, Urbana, IL, United States of America

⁵Yang Lou contributed equally to this work with Steven R Dolly and is considered as co-first author of this manuscript.

⁶Author to whom any correspondence should be addressed.

Abstract

In the majority of current radiation therapy (RT) applications, image quality is still assessed subjectively or by utilizing physical measures. A novel theory that applies objective task-based image quality assessment in radiation therapy (IQA-in-RT) was recently proposed, in which the area under the therapeutic operating characteristic curve (AUTOCT) was employed as the figure-of-merit (FOM) for evaluating RT effectiveness. Although theoretically more appealing than conventional subjective or physical measures, a comprehensive implementation and evaluation of this novel task-based IQA-in-RT theory is required for its further application in improving clinical RT. In this work, a practical and modular IQA-in-RT framework is presented for implementing this theory for the assessment of imaging components on the basis of RT treatment outcomes. Computer-simulation studies are conducted to demonstrate the feasibility and utility of the proposed IQA-in-RT framework in optimizing x-ray computed tomography (CT) pre-treatment imaging, including the optimization of CT imaging dose and image reconstruction parameters. The potential advantages of optimizing imaging components in the RT workflow by use of the AUTOCT as the FOM are also compared against those of other physical measures. The results demonstrate that optimization using the AUTOCT leads to selecting different parameters from those indicated by physical measures, potentially improving RT performance. The sources of systemic randomness and bias that affect the determination of the AUTOCT are also analyzed. The presented work provides a practical solution for the further investigation and analysis of the task-based IQA-in-RT theory and advances its applications in improving RT clinical practice and cancer patient care.

Keywords

task-based image quality assessment; radiation therapy; learning-based stochastic object models; geometric attribute distribution model; therapeutic operating characteristic curve

1. Introduction

Cancer remains the second leading cause of death globally. Despite the great clinical and pre-clinical efforts being spent on cancer research, the 5 year relative survival rate for all cancers diagnosed in 2004–2010 was 66.9%, with the lowest rate being only 7.7% (Siegel *et al* 2015). As a major cancer treatment procedure, radiation therapy (RT) is carried out on over half of the total number of cancer patients (Jaffray DA 2015). Optimal RT treatments seek to optimize tumor control while reducing damage to the surrounding healthy organs-at-risk (OARs), and ultimately to improve cancer patient survival rate and quality of life (Hoffmann *et al* 2013, Deasy *et al* 2015, Njeh *et al* 2015).

RT treatment heavily relies on the use of multi-modality imaging. Information extracted from the acquired patient images is utilized in all steps of the RT workflow, including tumor target delineation, OAR contouring, radiation treatment planning, treatment plan delivery, and treatment response assessment. Successful RT treatment relies on the performance of imaging systems and image processing/analysis algorithms. In current RT practice, optimization of the employed imaging systems and algorithms is still mainly guided by traditional physical measures of image quality such as noise level, spatial resolution, and image contrast. While being easy to use, these measures have certain limitations when employed to assess and optimize imaging systems in RT. First, they do not account for all sources of randomness in the measurement data (Barrett and Myers 2013). Second, they do not directly reflect the purpose and effectiveness of an image for use in RT, even though in diagnostic imaging it is widely accepted that imaging systems and algorithms should be optimized for their tasks by use of task-based image quality assessment (IQA) methodologies (Barrett *et al* 1995, 2010, Barrett *et al* 2013). Due to the different purposes of images employed in RT, there remains a great need to develop analogous task-based IQA methodologies to guide the optimization of imaging systems and algorithms for RT applications.

To address this need, a novel objective task-based IQA-in-RT theory has been proposed by Barrett *et al* (2010, 2013), Barrett *et al* (2015). In this IQA-in-RT theory, the therapy operating characteristic (TOC) curve (Moore and Mendelsohn 1972, Metz *et al* 1982) is employed, with the area under the TOC (AUTOC) being utilized as an innovative figure-of-merit (FOM) to measure the effectiveness of therapeutic outcomes and assess the performance of imaging systems and algorithms in RT. Theoretically, task-based IQA-in-RT is more appealing than conventional subjective or physical measures. However, this new theory must be comprehensively implemented and evaluated to achieve wide adoption by the RT community and potentially improve RT clinical practice. In order to accomplish this, several technical innovations are required but still remain undone.

In this study, a modular framework is proposed to address the need for a practical and comprehensive implementation of the novel task-based IQA-in-RT theory. The proposed IQA-in-RT framework implements each step of the radiation therapy planning process and the treatment outcome assessment. Specifically, stochastic object models (SOMs) are employed to accurately describe the variations in anatomy and tumor structures in specified ensembles of RT patients. The proposed modular computational framework can utilize the SOMs to compute the AUTO C and its ensemble average, which will serve as the FOMs for optimizing different imaging systems and algorithms that are employed in RT. The stochastic quantities that control the probabilities of tumor control and normal-tissue complications can also be analyzed and quantified with the implemented framework. This would permit the performance of imaging systems and algorithms to be optimized based on the assessment of therapeutic outcomes, the ultimate task of RT.

As an example demonstration, the IQA-in-RT framework is implemented and employed to optimize the image quality of pre-treatment computed tomography (CT) images (also referred to as CT simulation images) for external-beam prostate radiation treatment planning. In particular, the CT image reconstruction parameters are optimized to maximize treatment efficacy while minimizing the imaging dose. The results show that the AUTO C outperforms other conventional physical IQ assessment metrics for optimizing imaging components. The sources of systemic randomness and bias that affect the determination of AUTO C are also analyzed. The preliminary studies performed in this study demonstrate the applicability of the framework for RT task-based IQA and its superior performance compared to other physical IQ measures. The developed framework can potentially advance practical applications of the IQA-in-RT theory. Ultimately, it can guide the optimization of different components of the RT process based on treatment efficacy.

2. Background

2.1. Overview of the task-based IQA-in-RT theory

Radiation treatment seeks to optimize the radiation dose to the tumor while minimizing the radiation damage to OARs surrounding the tumor. Briefly speaking, the RT treatment workflow includes treatment planning and plan delivery. Imaging plays a critical role in each step of this workflow. Given an image dataset, the first step in RT planning is to estimate the boundaries of the tumor and nearby OARs that might be damaged by the treatment. The treatment plan strives to optimize the radiation dose within the tumor boundary while minimizing the dose within the boundaries of OARs. The plan delivery seeks to assure that the delivered radiation dose within tumor and OAR boundaries closely approximates the planned dose. The final treatment efficacy can be evaluated by two quantities: the tumor control probability (TCP) and the normal tissue complication probability (NTCP). TCP is the probability that radiation destroys the tumor, and NTCP is the probability of medical complications resulting from radiation damage to OARs. Both quantities are proportional to the radiation dose, meaning both will increase as the total radiation increases (Moore and Mendelsohn 1972, Tokars and Griem 1979, Hanley and McNeil 1982, Metz *et al* 1982, Andrews 1985, Ågren *et al* 1990, Hoffmann *et al* 2013).

According to the task-based IQA-in-RT theory proposed by Barrett *et al* (2010, 2013), Barrett *et al* (2015), the performance of radiation therapy can be evaluated by the TOC curve, a parametric plot of TCP versus NTCP. The area under the TOC curve (AUTO C) can be used as a FOM for assessing imaging and any other components that affect the RT, e.g. image acquisition, tumor/organ boundary estimation, treatment planning and plan delivery quality. An AUTO C of 1.0 represents an ideal treatment, and an AUTO C of 0.5 corresponds to a treatment scenario in which the probability of tumor control can be increased only at the expense of an identical increase in the probability of normal-tissue complications. The concepts of TOC and AUTO C in this IQA-in-RT theory are analogous to the receiver-operating-characteristic (ROC) curve and AUC in signal detection theory (Barrett and Myers 2013). By qualifying and analyzing TOC and AUTO C, the components within an RT workflow can be optimized for the specific task of maximizing radiation therapy treatment efficacy for an individual patient. In addition, this task-based metric can be averaged over a population of patients for the assessment and optimization of a particular component in the RT workflow.

2.2. Mathematical description of the task-based IQA-in-RT theory

This section reviews the salient mathematical aspects of the IQA-in-RT theory (Barrett *et al* 2010, 2013, Barrett *et al* 2015), focusing on the relation between imaging components within an RT planning workflow and TCP/ NTCP. It serves as a very brief introduction to the theories behind the motivation of task-based-IQA, which are described in more detail in a series of papers by Barrett *et al* (2010, 2013), Barrett *et al* (2015). A list of acronyms used in the paper is shown in table A3.

The randomness within the RT plan delivery process is neglected in this study. The treatment planning work-flow includes medical image acquisition, tumor/OAR segmentation, and treatment planning. Without loss of generality, in this study we assume that there is a group of J patients, $J > 1$. The set of organs of the j th patient is denoted as $\mathbf{O}_j$, and the true boundaries (continuous 3D surfaces) of these organs are denoted as $\mathbf{B}_j$. The image dataset acquired with a specific imaging system for the j th patient is denoted as $\mathbf{G}_j$.

The first step of RT planning is to delineate the boundaries of the tumor and surrounding OARs, given $\mathbf{G}_j$ as input. The true boundaries $\mathbf{B}_j$ are unknown in practice and only an estimated discrete representation of the boundaries $\hat{\mathbf{B}}_j$ based on $\mathbf{G}_j$ can be obtained from the acquired image dataset by use of a delineation algorithm. The second step of RT planning is to design a treatment plan based on the image set $\mathbf{G}_j$, the estimated boundaries $\hat{\mathbf{B}}_j$, and the pre-defined treatment prescription R_x . The treatment plan is created and adjusted to optimize the radiation dose delivered to the tumor while minimizing the radiation dose delivered to the surrounding OARs. The planned dose distribution to the patient can be denoted as $\mathbf{D}_j(\mathbf{r})$, where $\mathbf{r}$ is a position vector within the patient. Without loss of generality, $\mathbf{D}_j(\mathbf{r})$ can be represented as a scaled version of a reference unit dose $\mathbf{D}_{0,j}(\mathbf{r})$,

$$\mathbf{D}_j(\mathbf{r} \mid \lambda) = \lambda \mathbf{D}_{0,j}(\mathbf{r}), \quad (1)$$

where λ (a scale factor) and the unit dose $\mathbf{D}_0(\mathbf{r})$ can be chosen empirically based on medical experience. $\mathbf{D}_j(\mathbf{r} | \lambda)$ is also denoted $\mathbf{D}_j$ as for notational simplification with the dependence on implicit in this paper. The TCP and NTCP corresponding to a treatment plan can be mathematically denoted as $\mathbf{Pr}(TC)$ and $\mathbf{Pr}(NTC)$ respectively. The sources of randomness affecting $\mathbf{Pr}(TC)$ and $\mathbf{Pr}(NTC)$ can be modeled by conditional probability density functions (PDFs). The PDF in the obtained image set $\mathbf{G}_j$ for the j th patient can be denoted as $f(\mathbf{G}_j | j)$, which reflects the statistical nature of the imaging process. Given $\mathbf{G}_j$ the PDF for the estimated boundaries $\hat{\mathbf{B}}_j$ is denoted as $f(\hat{\mathbf{B}}_j | \mathbf{G}_j)$, which reflects the variability in the segmentation procedure, e.g. from manual segmentation or semi-automatic methods. The PDF for the actual dose distribution planned based on $\hat{\mathbf{B}}_j$ is expressed as $f(\mathbf{D}_j | \hat{\mathbf{B}}_j, \mathbf{G}_j, j)$, which also reflects the dose distribution variations caused by the randomness of dose calculation algorithms and other characteristics of the patient, such as fine details not captured in the image data. Therefore, $\mathbf{Pr}(TC)$ and $\mathbf{Pr}(NTC)$ depend on the actual dose distribution within the patient's body, on the true boundaries of tumor/OARs, and on other patient-dependent biological factors. These probabilities can be denoted as $\mathbf{Pr}(TC | \mathbf{D}_j, \mathbf{B}_j, j)$ and $\mathbf{Pr}(NTC | \mathbf{D}_j, \mathbf{B}_j, j)$ (Barrett *et al* 2013). By integrating over all possible $\mathbf{G}_j$, $\hat{\mathbf{B}}_j$, and $\mathbf{D}_j$ the probability of tumor control for the j th patient can be calculated as

$$\mathbf{Pr}(TC | \lambda, j) = \int \int \int \mathbf{Pr}(TC | \mathbf{D}_j, \mathbf{B}_j, j) f(\mathbf{D}_j | \hat{\mathbf{B}}_j, \mathbf{G}_j, \lambda, j) f(\hat{\mathbf{B}}_j | \mathbf{G}_j) f(\mathbf{G}_j | j) d\mathbf{D}_j d\hat{\mathbf{B}}_j d\mathbf{G}_j, \quad (2)$$

where the dependence on λ comes from equation (1). Similarly, the probability for normal tissue complication (NTCP) can be calculated as

$$\mathbf{Pr}(NTC | \lambda, j) = \int \int \int \mathbf{Pr}(NTC | \mathbf{D}_j, \mathbf{B}_j, j) f(\mathbf{D}_j | \hat{\mathbf{B}}_j, \mathbf{G}_j, \lambda, j) f(\hat{\mathbf{B}}_j | \mathbf{G}_j) f(\mathbf{G}_j | j) d\mathbf{D}_j d\hat{\mathbf{B}}_j d\mathbf{G}_j.$$

Theoretically, equations (2) and (3) can be employed to describe and assess imaging components within the RT workflow, including imaging acquisition techniques, image reconstruction algorithms, tumor/OAR delineation methods, treatment planning methods, and treatment dose calculation algorithms. When implementing the IQA-in-RT theory for assessing these components within a RT workflow, the following two assumptions will be employed:

1. First, the dose calculation algorithm can be assumed to be accurate, non-random, and independent of the patient, meaning that $f(\mathbf{D}_j | \hat{\mathbf{B}}_j, \mathbf{G}_j, j)$ can be approximated by a Dirac delta distribution,

$$f(\mathbf{D}_j | \hat{\mathbf{B}}_j, \mathbf{G}_j, \lambda, j) = \delta[\mathbf{D}_j - \mathbf{D}(\hat{\mathbf{B}}_j, \mathbf{G}_j | \lambda)]. \quad (4)$$

Here $\mathbf{D}(\hat{\mathbf{B}}_j, \mathbf{G}_j | \lambda)$ represents the dose distribution calculated by the planning algorithm (Barrett *et al* 2013). $\oint[\cdot]$ can be used to perform the integral over $\mathbf{D}_j$ therefore, equation (2) can be simplified to

$$Pr(TC | \lambda, j) = \int \int Pr(TC | D(\hat{\mathbf{B}}_j, \mathbf{G}_j | \lambda), \mathbf{B}_j, j) f(\hat{\mathbf{B}}_j | \mathbf{G}_j) f(\mathbf{G}_j | j) d\hat{\mathbf{B}}_j d\mathbf{G}_j. \quad (5)$$

2. Second, we can neglect the patient-to-patient variation in tissue response to ionizing radiation and have

$$Pr(TC | D_j, \mathbf{B}_j, j) = Pr(TC | D_j, \mathbf{B}_j). \quad (6)$$

Based on the above, and noting that $f(\hat{\mathbf{B}}_j | \mathbf{G}_j) f(\mathbf{G}_j | j) = f(\hat{\mathbf{B}}_j, \mathbf{G}_j | j)$, the TCP and NTCP expressions in equations (2) and (3) can be simplified to

$$Pr(TC | \lambda, j) = \int \int Pr(TC | D(\hat{\mathbf{B}}_j, \mathbf{G}_j | \lambda), \mathbf{B}_j) f(\hat{\mathbf{B}}_j, \mathbf{G}_j | j) d\hat{\mathbf{B}}_j d\mathbf{G}_j, \quad (7)$$

$$Pr(NTC | \lambda, j) = \int \int Pr(NTC | D(\hat{\mathbf{B}}_j, \mathbf{G}_j | \lambda), \mathbf{B}_j) f(\hat{\mathbf{B}}_j, \mathbf{G}_j | j) d\hat{\mathbf{B}}_j d\mathbf{G}_j. \quad (8)$$

Empirical estimation of the TCP/NTCP values in equations (7) and (8) can be obtained from the K generated random image samples of the j th patient. For a given j th patient with known tumor/OAR boundaries $\mathbf{B}_j$, a set of images $\mathbf{G}_j^k$, ($k \in \{1, \dots, K\}$) can be generated directly, which simulate the noisy measurement data recorded by the imaging systems and the employed image reconstruction process. Then contour delineation can be performed to estimate the organ boundaries $\hat{\mathbf{B}}_j^k$ from $\mathbf{G}_j^k$. From the law of large numbers (Rice 2006), when only one delineated organ set is obtained for each image set and K is large enough, equations (7) and (8) can be approximated as

$$\widehat{Pr}(TC | \lambda, j) = \frac{1}{K} \sum_{k=1}^K Pr(TC | D(\hat{\mathbf{B}}_j^k, \mathbf{G}_j^k | \lambda), \mathbf{B}_j), \quad (9)$$

$$\widehat{Pr}(NTC | \lambda, j) = \frac{1}{K} \sum_{k=1}^K Pr(NTC | D(\hat{\mathbf{B}}_j^k, \mathbf{G}_j^k | \lambda), \mathbf{B}_j). \quad (10)$$

Finally, by averaging the obtained $\widehat{Pr}(TC | \lambda, j)$ and $\widehat{Pr}(NTC | \lambda, j)$ over an ensemble of J patients, these two quantities can be empirically estimated over a patient population

$$\widehat{Pr}(TC | \lambda) = \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K Pr(TC | D(\hat{\mathbf{B}}_j^k, \mathbf{G}_j^k | \lambda), \mathbf{B}_j), \quad (11)$$

$$\widehat{Pr}(NTC | \lambda) = \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K Pr(NTC | \mathbf{D}(\hat{B}_j^k, \mathbf{G}_j^k | \lambda), \mathbf{B}_j). \quad (12)$$

A TOC curve can be constructed by changing the absolute dose level in the RT treatment, which is equivalent to varying λ in equations (11) and (12). Therefore, the AUTO C can be employed as the final FOM to quantify therapy efficacy and assess the performance of imaging components in the RT workflow.

3. Implementation of the IQA-in-RT theory

In this study, a semi-automated framework is proposed for implementing the IQA-in-RT theory. Figure 1 outlines the proposed framework, which includes three inputs, three major processing modules, and one output. Each input and processing module of this framework can be implemented with different algorithms and processes, according to specific user needs and RT tasks. The framework can be employed to investigate how the theory assists and improves the clinical RT practices and facilitate its application in RT. The simulation framework implemented here focuses on the imaging components in the RT planning workflow.

The first input of the IQA-in-RT framework is a stochastic object model (SOM) (Dolly *et al* 2018). When optimizing an imaging system by use of task-based measures of image quality via computer-simulation, it is important to account for as many sources of randomness in the measurement data as possible. Variations in the to-be-imaged object are one important source of randomness. The SOM represents this randomness by capturing the statistical variations in human anatomy relevant to a specific tumor site within an ensemble of patients. The second input is an imaging model for generating a collection of images directly from the SOM. Ideally, these images should reflect all sources of randomness in the imaging system. The third input consists of information related to the RT tasks, including the prescribed dose to the target, dose constraints for surrounding OARs, and details related to the treatment modality (e.g. 3D conformal radiation therapy, intensity-modulated radiation therapy (IMRT), brachytherapy, etc).

In the first processing module, the target tumor and surrounding OARs are segmented (or contoured) from the simulated images. In the second processing module, an RT treatment plan is created and optimized based on the pre-defined treatment prescription and the estimated boundaries. In the third processing module, the treatment is evaluated by estimating the patient treatment outcome based on the optimized dose distribution and the known ground truth organ boundaries. This module employs a radiobiological model using the treatment dose to the actual tumor/OARs to compute TCP and NTCP based on equations (7) and (8).

Finally, given the three inputs mentioned earlier and the three processing modules just described, the output of the framework is the calculated AUTO C value. The AUTO C is used as the FOM for assessing and optimizing imaging components within the RT workflow.

4. Example implementation of the proposed IQA-in-RT framework

In this section, an IMRT plan for human prostate cancer treatment based on x-ray CT imaging is presented as an example implementation of the IQA-in-RT framework described in section 3. The entire framework is implemented in C++ to enable extendability with other libraries, including ITK (Ibanez *et al* 2005), FFTW (Frigo and Johnson 1998), and OpenMP (Gabriel *et al* 2004). Packages written in other programming languages are also incorporated via a wrapper class interface.

4.1. Input 1: stochastic object model

The first input of the framework is an established stochastic object model (SOM). A novel and computationally tractable methodology was employed to learn a SOM from a set of high-quality volumetric clinical images with segmented organ contours (Dolly *et al* 2018). Figure 2 shows the general workflow for learning such a SOM. The procedure includes five steps: (1) collecting patient data, (2) learning the multi-organ centroid geometric attribute distribution (GAD) that describe the inter-organ position distributions, (3) learning individual organ shape GADs that depicts the shape variations of specific organs among different patients, (4) generating random organs by sampling the learned GADs, and (5) creating an ensemble of numerical phantoms from the random organs. In this way, an ensemble of numerical phantoms can be assembled, whose anatomic variations mimic those of the patient population utilized in generating the GADs, and can satisfy the requirements of the task-based measures of image quality.

In this example implementation, we established an ensemble of human pelvis phantoms, including the prostate, bladder, rectum, femoral heads, pelvic bone, and seminal vesicles. The SOM has two purposes in the IQA-in-RT framework. First, the generated numerical phantoms are employed to simulate images that can be used for target/OAR contouring and RT planning. Second, the phantom provides ground truth target/OAR contours, which are employed to calculate the TCP and NTCP in the third processing module. section 4.6 describes this process.

4.2. Input 2: imaging model for simulating images based on SOMs

In this subsection, the process of generating CT simulation images by use of the established pelvis phantom is briefly summarized, including the simulation of x-ray projection measurements and reconstruction of CT images from the measurement data. A full mathematical description was presented in a previous publication (Dolly *et al* 2018).

4.2.1. CT projection data generation—To generate CT images of a given numerical pelvis phantom, projection measurement data were first simulated by tracing a ray from a poly-energetic x-ray source through the numerical phantom to each detector element, then calculating the intensity reduction according to exponential attenuation. The initial source intensity was set by a user-adjustable parameter I_0 , which represents the number of photons originating from each source ray. As I_0 increases, the noise in the projection domain decreases, but the imaging radiation dose to the patient increases. The discrete poly-energetic imaging source spectrum was generated by the TASMIP algorithm (Boone and

Seibert 1997), resulting in a spectrum sampled at integer kilo Voltage peak (kVp) intervals from 1–150 kV. The properties of each organ in the numerical phantom were assigned according to the NIST database online (US Department of Commerce 2009), and are shown in table A1 in appendix A.1 .

The organs within the numerical phantoms were represented as triangular meshes, so the pathlength of the CT source rays through each organ/material was calculated using the highly optimized Moller–Trumbore ray-triangle intersection algorithm (Möller and Trumbore 2005). Prior to image reconstruction, Poisson noise was added to the measured data. Finally, a Gaussian blur was applied to the projection data to simulate the degraded spatial resolution from signal spread within CT detectors.

4.2.2. CT image reconstruction—To accommodate multi-slice, helically-acquired projection data, a filtered interpolation approach (Taguchi and Aradate 1998) in the z -dimension was utilized to obtain single-slice, axial projection data. The filtered interpolation approach includes two steps: (1) resampling by linear interpolation using adjacent data points, and (2) filtering the resampled dataset. The acquired measurement data was filtered with a Fourier-domain based filter that was applied prior to back-projection. The axial slices were then reconstructed using a standard filtered backprojection approach for fan-beam data (Kak and Slaney 2001). A sinc filter was utilized with an user-adjustable parameter α to allow for a trade-off between image sharpness and noise reduction. The parameter α can be chosen according to the desired imaging aims. As α increases, the image noise decreases but so does the image sharpness. Post-reconstruction CT images were saved in DICOM format using the ITK C++ library (Yoo *et al* 2002). Examples of the reconstructed CT images are shown in figures 4 and 8, which also respectively show the effect of I_0 and α on the reconstructed image appearance.

4.3. Input 3: treatment prescription

The treatment prescription is used as a criterion to guide the treatment planning process, and varies for different treatment sites and procedures. Recommended treatment prescriptions typically require information about the (1) treatment modality, and (2) relevant dose statistics such as minimum/maximum/mean dose, and dose volume histogram (DVH) (Drzymala *et al* 1991) constraints for both tumors/targets and OARs. The treatment prescription for an IMRT prostate treatment is described in table 1 below.

4.4. Module 1: image segmentation

In this first processing module, the simulated prostate CT images in section 4.2 are imported into an open source software, ITK-SNAP (Yushkevich *et al* 2006), for segmenting organ contours, which are later used in section 4.5 for treatment planning. The ITK-SNAP software combines active contour methods with manual delineation and image navigation. Segmentation in ITK-SNAP is comprised of three steps. First, a 3D rectangular prism region of interest (ROI) that encloses the organ to-be-segmented is manually selected using the image navigation features. Second, a pre-segmented speed image is created. A thresholding algorithm, which maps the voxel intensities within a specified range to positive speed values, is utilized to create the speed image, resulting in pixels inside the organ having values close

to 1, and voxels outside having values close to 0. Finally, an active contour method (Caselles *et al* 1997) is applied to the speed image to obtain the final segmentation results. All parameters in the speed-image-creation and active contour method are set as constant values to maintain consistency of segmentation results across all images. Once the segmentation has been performed, the segmented structures are exported as triangular meshes for treatment planning and dose calculation.

4.5. Module 2: radiation treatment planning

Next the IMRT planning for prostate cancer treatment is implemented, which includes the three steps of treatment plan initialization, dose calculation, and dose optimization.

4.5.1. Treatment plan initialization Five nearly-equiangular beams were utilized at beam angles of—35°, 100°, 180°, 270°, and 325°, with all beams directed towards a common treatment isocenter, and with a source-axis-distance (SAD) of 100 cm (Khan and Gerbi 2012). Each beam was split into a grid of 20×20 square beamlets, each having a side length of 1 cm at the isocenter plane. The treatment isocenter was chosen to be the centroid of the segmented prostate, which can be automatically calculated without human interaction, thus preventing human bias in isocenter placement.

Table 1 above summarized the treatment prescription and dose-volume histogram (DVH) constraints of target/OARs used to optimize the dose distribution. The dose prescriptions and constraints for prostate IMRT treatment were obtained from a literature review (Khan and Gerbi 2012). Prior to dose calculation, each beamlet was checked to determine if it passes through the target organ (i.e. prostate), and was marked as open if so, and closed otherwise.

4.5.2. Treatment dose calculation—An open-source Monte Carlo dose calculation engine, dose planning machine (DPM) (Sempau *et al* 2000), was utilized for treatment dose calculation. The original DPM source code was adjusted in this study to accommodate multiple beamlets, as in an IMRT treatment. Given the isocenter, beam angle, and beamlet number (i.e. $0 \rightarrow 0\ 399$ for a 20×20 beamlet grid) as inputs from treatment planning, the adjusted DPM program calculated the radiation beamlet incidence on the 3D CT image set. Six different materials were considered: air, lung, adipose tissue, soft tissue, skeletal muscle, and cortical bone (Vanderstraeten *et al* 2007). The CT-number-to-material conversion was then performed based on the linear attenuation coefficients from the NIST database. The source energy spectrum for a 20 MV photon beam was used, and the number of photons per beamlet set to 1.5×10^6 .

4.5.3. Treatment dose optimization—The initial dose calculated in section 4.5.2 assumes uniform beamlet weights, meaning each beamlet emits the same number of photons. However, to better satisfy the dose constraints, the beamlet weights should be adjusted to yield an optimal dose distribution with respect to the treatment prescription, which is the aim of this treatment dose optimization step.

A simulated-annealing-based algorithm was implemented for dose optimization (Webb 1991, Shepard *et al* 2002). The weights of beamlets are iteratively changed to improve dose

distribution by minimizing a cost function F . Denote the weight of the b th beamlet at the i th iteration as ω_b^i , $b \in B$

. The dose distribution $\mathbf{D}^i(\mathbf{r})$ at location $\mathbf{r}$ in the i th iteration is then computed as

$$\mathbf{D}^i(\mathbf{r}) = \sum_{b=1}^B \omega_b^i d_b(\mathbf{r}), \quad (13)$$

where $d_b(\mathbf{r})$ denotes the calculated dose distribution of the b th beamlet at location $\mathbf{r}$ by use of DPM. Discretizing equation (13) yields

$$\mathbf{D}^i = \sum_{b=1}^B \omega_b^i \mathbf{d}_b, \quad (14)$$

where $\mathbf{D}^i$ and $\mathbf{d}_b$ denote $D(\mathbf{r})$ and $d_b(\mathbf{r})$ evaluated on a voxelized grid and vectorized. Therefore, the cost function F is defined as

$$F(\mathbf{D}^i) = \sum_{n=1}^{N_c} v_n (C_n(\mathbf{D}^i) - C_n^{ideal})^2, \quad (15)$$

where each dose constraint C_n , $n = 1, 2, \dots, N_c$, is a function that maps $\mathbf{D}^i$ to a real scalar, and has an idealized value C_n^{ideal} if the constraint is fully satisfied. F evaluates how closely the current achieved dose constraints $\{C_n\}$ satisfy the pre-defined dose constraints C_n^{ideal} .

The weight, v_n , for each dose constraint represents the priority of a constraint that a treatment plan needs to satisfy. The weights are normally determined by physicians based on their clinical experience and other clinical judgments. C_n^{ideal} can be set based on the minimum dose, maximum dose, or by a point on the dose volume histogram curve. The dose constraints used in our study are listed in table 1, with $N_c = 7$. The weights are set to 10, 5, 5, and 3 for prostate, bladder, rectum, and femur, respectively.

We denote the weights of all beamlets in the i th iteration as a weight vector

$\mathbf{W}^i = \{\omega_1^i, \dots, \omega_B^i\}$, which we use in the simulated annealing algorithm described below.

1. Initialize beamlet weights to unity: $\mathbf{W}^0 = \mathbf{1}$.
2. At the i th iteration, randomly pick one beamlet weight ω_b^i , and adjust its weight to $\tilde{\omega}_b^i$ by adding a pre-determined perturbation amount g , with its sign being chosen as positive or negative with equal probability,

$$\tilde{\omega}_b^i = \omega_b^i \pm g. \quad (16)$$

3. Calculate the dose distribution associated with $\mathbf{W}^i$ and $\tilde{\mathbf{W}}^i$, denoted as $\mathbf{D}^i$ and $\tilde{\mathbf{D}}^i$, respectively.

4. Compute the cost functions, $F(\mathbf{D}^i)$ and $F(\tilde{\mathbf{D}}^i)$ before and after the adjustment according to equation (15).
 - a. If $F(\tilde{\mathbf{D}}^i) < F(\mathbf{D}^i)$, replace $\mathbf{W}^i$ with $\tilde{\mathbf{W}}^i$;
 - b. If $F(\tilde{\mathbf{D}}^i) \geq F(\mathbf{D}^i)$, the beamlet acceptance is defined by a probability P based on a Boltzmann simulated annealing cooling schedule (Shepard *et al* 2002). P is defined as,

$$P = 2A \times (1 + \exp(\log(n_{accept} + 1)/T_0))^{-1}, \quad (17)$$

where A and T_0 are pre-defined parameters. P decreases as the number of previously accepted adjustments (denoted by n_{accept}) increases.

5. Terminate the procedure after the number of accepted changes n_{accept} reaches a user-defined maximum n_{max} .

To avoid negative or non-realistic beamlet weights, a change is always rejected if it would result in a non-realistic weight. In our study, values for g , A , T_0 , and n_{max} were empirically chosen to be 0.1, 0.5, 1.0, and 1000, respectively, to produce adequate dose distributions while keeping the calculation time reasonable. A simple visualization program was created using the OpenGL graphics API to view the optimized dose distribution overlaid on the CT images (figure 5 below shows an example dose distribution). The final optimized beamlet weights and dose distribution were employed for calculating TCP and NTCP in the third processing module, as described in section 4.6 below.

4.6. Module 3: radiobiological modeling

After the treatment planning is completed, the treatment performance is evaluated in the third processing module. The evaluation can be performed by estimating the patient treatment outcome based on the optimized dose distribution and the known ground truth boundaries. Here, the equivalent uniform dose (EUD) (Niemierko 1997) within a specific organ or tumor of interest is implemented as an example radiobiological model. The EUD for either an organ or tumor is defined as

$$\text{EUD} = \left(\sum_{v=1}^V \kappa_v [\mathbf{D}]_v^{\alpha_c} \right)^{\frac{1}{\alpha_c}}, \quad (18)$$

where V is the total number of voxels within the delineated target or organ volume, κ_v is a binary value that equals 1 if the v th voxel falls within the current organ, and equals 0 otherwise, and $[\mathbf{D}]_v$ denotes the value at the v th voxel of the optimized dose distribution obtained from the second processing module. The term α_c is a tissue-specific parameter which describes the dose-volume effect. The term α_c is negative for tumors and positive for normal organs. Given the EUD of a target or organ, the related TCP or NTCP is calculated as (Niemierko 1997, Gay and Niemierko 2007)

$$TCP = \frac{1}{1 + \left(\frac{TCD_{50}}{EUD_{tumor}} \right)^{4\gamma_{50}}}, \quad NTCP_{organ} = \frac{1}{1 + \left(\frac{TD_{50,organ}}{EUD_{organ}} \right)^{4\gamma_{50}}}, \quad (19)$$

where TCD_{50} is defined as the radiation dose to control 50% of the tumor when the tumor is homogeneously irradiated, $TD_{50,organ}$ is defined as the tolerance dose for a 50% complication rate at a specific time interval (e.g. 5 years) when the current organ is homogeneously irradiated, and γ_{50} is a unitless model parameter that describes the slope of the dose response curve, and is the same for both tumor and normal organs.

The values of the parameters α_e , and TCD_{50} , TD_{50} , γ_{50} for all relevant tissues were obtained from the literature to calculate the TCP and NTCP (Emami *et al* 1991, Rana and Cheng 2014), and are shown in table A2 appendix A.2. The EUD is affected by the dose scalar parameter λ according to equations (11) and (12). To compute the therapeutic operating characteristic (TOC) curve, a range of λ values were swept, and for each value λ of , new EUD, TCP, and NTCP values were calculated, producing a new point on the TOC curve. The range of λ values was chosen so that the TCP and NTCP were fully sampled from 0.0001 to 0.9999. By numerically integrating the TOC curve using the trapezoid rule, the AUTO C value could be calculated. The NTCP curve was generated for each OAR in a similar way, and the corresponding AUTO C value represents the risk of damaging the current organ with respect to the tumor target. By repeating this procedure, an AUTO C value was obtained for each organ. Since the TCP is included in each AUTO C and tumor control is desired in all cases, multiplying the AUTO C values for all N_s OARs serves as a suitable cumulative AUTO C,

$$AUTO C_{cum} = \prod_{s=1}^{N_s} AUTO C_s. \quad (20)$$

This value serves as the final FOM, and summarizes the assessment of image quality, and therefore the patient treatment outcome.

5. Case studies that utilize the implemented framework

In sections 5.1 and 5.2, the feasibility and performance of the implemented IQA-in-RT framework for prostate cancer treatment are demonstrated by presenting its applications in two scenarios, for the optimization of (1) the CT imaging dose and (2) the CT image reconstruction parameter. Both optimizations improve the treatment efficacy of IMRT prostate treatments. In addition, in section 5.3, the AUTO C is compared to other IQA metrics to demonstrate an improved performance when using AUTO C as the FOM for the two scenarios. Finally, in sections 5.4 and 5.5, the performance of the IQA-in-RT framework is analyzed for sources of bias which can arise from different radiobiological models, and for sources of calculation randomness within the simulated annealing dose optimization algorithm.

5.1. Application 1: optimization of CT imaging dose

The imaging-related radiation dose has raised increasing concerns in RT with the rapid usage of x-ray based image guidance (i.e. IGRT) in recent years (Yan *et al* 1997, Hall and Wu 2003, Dawson and Jaffray 2007, Otto 2008). In particular, young patients risk more secondary cancers. In the first example application, we employed the proposed IQA-in-RT framework to minimize the CT imaging dose while maintaining the efficacy of the IMRT prostate cancer treatment.

A pelvis SOM with known organs was first created based on the methodology described in section 4.1 (Dolly *et al* 2018). Anatomically realistic numerical representations of organs with different shapes and positions were generated in a flexible and realistic manner by randomly sampling the established SOM. Figure 3(a) displays the interior of an example realization of the established SOM, including the prostate, bladder, rectum, femoral head, pelvic bone, and seminal vesicles. A 2D transverse cross-section of the phantom with external body contour is also shown in figure 3(b).

The purpose of this application is to optimize imaging dose ($CTDI_{vol}$) with constant reconstruction filter parameter (α). We chose a constant $\alpha = 2$ since this is the minimum amount of filter smoothing that produced acceptable images for all values of $CTDI_{vol}$. By acceptable images we mean that there are no severe artifacts, such as photon starvation or beam hardening, which would have made the images completely unusable for treatment. Since both the filter parameter and the imaging dose affect image noise as mentioned above, choosing the minimum amount of filter smoothing most clearly emphasizes the impact of imaging dose on image quality. By altering kVp and the number of initial photons (the variable I_0 , from section 4.2.1) of the imaging source, a set of CT images were simulated with varied imaging doses. To compare different imaging protocols, a relative volumetric CT dose index () was utilized to calibrate the CT acquisition process,

$$CTDI_{vol}(i, j) = \frac{CTDI_{val, i, j}}{CTDI_{vol, min}} = \left(\frac{I_i}{I_{min}} \right) \times \left(\frac{kVp_j}{kVp_{min}} \right)^2. \quad (21)$$

Here, the index (i, j) represents the i th setting of photon number I_0 and the j th kVp setting, both normalized by the minimal photon fluence I_{min} and the minimal kVp setting kVp_{min} . Eight settings of photon numbers were simulated as $I_0 = \{5 \times 10^3, 7.5 \times 10^3, 1.0 \times 10^4, 2.5 \times 10^4, 5.0 \times 10^4, 1.0 \times 10^5 \text{ and } 2.5 \times 10^4\}$, among which $I_{min} = 5 \times 10^3$. Similarly, four kVp settings were chosen as $kVp = \{80, 100, 120, 140\}$ among which $kVp_{min} = 80$. For each relative $CTDI_{vol}$ setting, 1000 image sets were used, with $J = 20$ patients and $K = 50$ images per patient, as in equations (11) and (12). Figure 4 displays example CT images corresponding to the same phantom but with varied photon numbers (representing varied imaging doses).

For each generated CT image, the target (prostate in this study) and organs were segmented by use of the method described in section 4.4. Then the related treatment plan was generated and optimized based on the image and the segmented organ contours. Figure 5 displays the image along with the dose distributions and DVHs (Drzymala *et al* 1991) for an example optimized plan based on a CT image set and the segmented target/OAR contours.

The TCP, NTCP, and TOC curves were calculated for each treatment plan and with the ground truth contours of target/OARs. The choice of normal tissue complication in this instance is rectal bleeding, the most common adverse effect from radiation therapy of prostate cancer. Figure 6 displays example TCP, NTCP, and TOC curves calculated for a treatment plan.

Finally, the AUTO C was calculated for each TOC curve, and the averaged AUTO C over the patient population was plotted as a function of the relative $CTDI_{vol}$ in figure 7. It can be seen that when the relative $CTDI_{vol} < 1.5$, the treatment plan efficacy is unacceptable because the AUTO C value is close to 0. As the imaging dose increases, the AUTO C rises to values close to 1, indicating good treatment plan efficacy. However, when the relative $CTDI_{vol} > 1.5$, the additional improvement in treatment plan efficacy becomes minimal. Therefore, for this particular application, it is desirable to set the $CTDI_{vol}$ relative value close to $CTDI_{vol} = 1.5$.

5.2. Application 2: optimization of CT image reconstruction filter

In the second example, the IQA-in-RT framework was applied to optimize the filter parameter α in the filtered back-projection (FBP) reconstruction algorithm (Dolly *et al* 2018). The purpose was to optimize α with constant imaging dose. We chose a constant number of photons per detector of $I_0 = I_{min} = 5 \times 10^3$ since this is the minimum amount of photons per detector that produced acceptable images for all values of α . Choosing the minimum amount of I_0 most clearly emphasizes the impact of the parameter of α on image quality. The setting $I_0 = 5 \times 10^3$ and $kVp = 140$ was used to minimize the photon starvation artifacts. The resulting value for $CTDI_{vol}$ was 3.05. We analyzed the trade-off between image noise and spatial resolution, as well as its impact on the radiation therapy treatment efficacy. Specifically, we swept α across a range of values: 2, 4, 6, 8, and 10. The ramp filter implementation corresponding to $\alpha \approx 0$ was excluded because the produced low-dose images were too noisy for adequate segmentation and treatment planning. For each α , 1000 images were created with $J = 20$ and $K = 50$, as shown in equations (11) and (12). kVp was chosen to be 140, and the number of photons per detector element was set to 5×10^3 , which corresponds to a relative $CTDI_{vol} = 3.05$. Examples of the simulated CT images reconstructed with various filter parameters are shown in figure 8.

Following the steps similar to those shown in section 5.1, the AUTO C values were calculated and averaged over all 1000 images for each α . Figure 9 displays the AUTO C versus the filter parameter α . It can be seen that images reconstructed using $\alpha = 2$ are not suitable for patient treatment. The AUTO C then increases as α increases, reaching a maximum at $\alpha = 6$. Finally, due to the over-smoothing on images reconstructed with even higher α values, the AUTO C decreases, albeit only slightly. These results indicate that the proposed framework can be used to objectively select an α value for the FBP reconstruction filter which optimally maximizes patient treatment efficacy.

5.3. Analysis 1: comparison of AUTO C to other IQA metrics

To evaluate the performance of the AUTO C as a FOM for the IQA-in-RT framework, we compared the AUTO C with other commonly used physical metrics for image quality assessment. The compared metrics include the commonly-employed contrast-to-noise ratio

(CNR), structure similarity (SSIM) index, and noise equivalent quanta (NEQ). In addition, since segmentation accuracy plays an important role in treatment plan quality (Li *et al* 2014), Dice's similarity coefficient (DSC) between the segmented organ contours and known organ contours is commonly used as an indicator of segmentation accuracy. Therefore, DSC served as an additional physical IQA metric in the comparison. The CNR was calculated by selecting target organ (prostate) regions of interest (ROIs) in the generated images. The SSIM was calculated for each of the images according to Wang *et al* (2004). Both the CNR and SSIM were then averaged over all images in a patient dataset. To calculate the noise equivalent quanta (NEQ), the modulation transfer function (MTF) and the noise power spectrum (NPS) were first calculated separately, and NEQ was calculated by use of the approach implemented by Friedman *et al* (2013). Details of these calculations are shown in our previous work (Dolly *et al* 2016).

The four aforementioned IQA metrics, along with the AUTO C, were calculated with respect to both the relative CTDI_{vol} value in section 5.1 and the FBP parameter α in section 5.2, and are plotted in figure 10. As shown in figure 10(a), while the calculated IQA metrics all share the same general trend, the AUTO C implies a clear threshold for a minimum imaging dose that maintains treatment plan quality. Figure 10(b) shows that different IQA metrics lead to different optimal values of the reconstruction filter parameter α . When image quality is being assessed based on CNR, NEQ, or DSC, the most optimal reconstructions are obtained when α is equal to 10, 10, or 8, respectively. Comparatively, the optimal α is equal to 6 when using AUTO C as the FOM. Therefore, it can be seen that different optimal values for α are obtained when different IQA metrics are used as FOMs. It is worth emphasizing here that the meaning of optimal in these cases depends on the metric and desired task. For example, to maximize segmentation accuracy, the optimal reconstruction occurs when $\alpha = 8$, where the DSC is at a maximum. However, the ultimate task of radiation therapy is to optimize treatment efficacy instead of target/OAR segmentation accuracy.

5.4. Analysis 2: sources of bias

The AUTO C value directly depends on the TCP and NTCP, which are calculated from pre-defined radiobiological models. To investigate the bias introduced by different radiobiological models, we employed two other models (Warkentin *et al* 2004) to calculate the TCP and NTCP, and compared the resulting AUTO C to that calculated based on the EUD model in equation (19).

First, the TCP was alternatively calculated using a Poisson-based model (Webb and Nahum 1993),

$$\text{TCP} = \left(\frac{1}{2}\right)^x, \quad x = \sum_i v_i \times \exp\left(2\gamma_{50} \frac{(1 - D_i/\text{TCD}_{50})}{\ln 2}\right), \quad (22)$$

where $\{D_i, v_i\}$ represents a dose-volume pair of the differential dose-volume histogram (DDVH).

Second, the NTCP of a specific organ was alternatively calculated using a sigmoidal dose response (SDR) model (Lyman 1985),

$$\begin{aligned} \text{NTCP} &= \Phi\left(\frac{\text{EUD}-\text{TD}_{50, \text{organ}}}{\gamma_{50}\text{TD}_{50, \text{organ}}}\right), \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt \\ &= \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right], \end{aligned} \quad (23)$$

where $\Phi(x)$ is the probit function. The EUD is equal to a generalized mean dose (GMD), calculated from the dose-volume pairs D_i, v_i of the DDVH, based on equation (18). Together with the EUD-based TCP and NTCP calculations, four possible combinations of TCP-NTCP calculations can be created: EUD-EUD, Poisson-EUD, EUD-SDR, and Poisson-SDR. The AUTOCs were calculated for all four combinations by varying the dose parameter CTDI_{vol} and the FBP reconstruction parameter α . In addition to the choice of model type, it is also known that the TCP/NTCP calculation is sensitive to parameters used within the model, such as γ_{50} , TCD_{50} , and $\text{TD}_{50, \text{organ}}$. For example, γ_{50} and TCD_{50} for tumors vary based on the progression staging of the cancer (Okunieff *et al* 1995). Hence, to study the bias of these parameters, AUTOC values were calculated for various γ_{50} and TCD_{50} values corresponding to different prostate cancer stages (T0–T4) (Okunieff *et al* 1995). These AUTOC values were calculated with varying CTDI_{vol} and α , as in sections 5.1 and 5.2.

Figure 11 displays the effect on the AUTOC of using different radiobiological models and staging parameters to calculate TCP or NTCP. Using EUD yields better AUTOC values than the other three combinations. However, all four AUTOC curves correspond to the same optimal reconstruction filter ($\alpha = 6$), as shown in figures 11(b) and (d), and the same relative CTDI_{vol} for optimizing imaging dose, as shown in figures 11(a) and (c).

5.5. Analysis 3: sources of randomness

The initial derivation of the TOC curve, as described in section 2.2 and summarized by equation (9), is predicated on a non-random dose calculation algorithm. However, as stated in section 4.5, the simulated annealing algorithm utilizes a random number generator to select the beamlet number and the sign of the pre-determined perturbation amount g . In order to study the randomness introduced into the AUTOC calculation, the simulated annealing algorithm was repeated 100 times for each image dataset to obtain a sufficient sample size, and the subsequent AUTOC for each optimization was calculated and analyzed. In this way, the AUTOC variation introduced by the dose optimization algorithm was found to be approximately 0.02%, at the 95% confidence interval. This finding implies that the proposed framework is insensitive to the randomness and bias of the simulated-annealing algorithm.

6. Discussion

Intermediate results in RT, such as the agreement between ground-truth and auto-segmentation contours and CT numbers, can represent the direct impact of various imaging parameters (such as scanning dose) on the treatment process. However, the most important impact in radiation therapy is the overall treatment efficacy, of which segmentation or CT number accuracy are intermediate steps of the RT process. Therefore, a metric that quantifies treatment efficiency, such as the AUTOC, should be employed to study the ultimate impact

of imaging parameter selection on patient care. We proposed and implemented a modular computational framework for computing the AUTOC for RT workflows that involve imaging components. The modular design of the framework is easy to use to validate a specific component within the RT planning workflow, based on the ultimate task of RT treatment efficacy. The proposed IQA-in-RT framework enables an assessment of not only the direct impact, but also how that direct impact is propagated through the entire treatment process. Moreover, the AUTOC quantifies many effects of image quality (including segmentation and CT number accuracy) into a single value, making it a comprehensive metric.

As shown in figure 10, the AUTOC produces different results compared to the DSC, which demonstrates that the segmentation accuracy is not the only relevant effect of image quality on treatment efficacy. Also, figure 10(a) demonstrates that the optimal CT imaging dose determined by using the AUTOC is lower than that determined using the DSC. This signifies that segmentation accuracy is only a part of the RT process and optimizing imaging parameters solely to improve the DSC may have sub-optimal treatment efficacy, which should be the first priority. We have observed that the SSIM performs similarly to AUTOC for the example shown in figure 10. In particular, figure 10(b) yields the same optimal reconstruction parameter ($\alpha = 6$) for both AUTOC and SSIM. However, in figure 10(a) it is apparent that the AUTOC curve has a clearly defined cut-off point at a particular imaging dose, while the SSIM decreases gradually as the imaging dose decreases. Practically, this means that a clear optimal imaging dose can be selected when using the AUTOC, while the dose would be more arbitrarily chosen for the SSIM. This confirms the uniqueness of the AUTOC. It is possible that metrics like the SSIM might perform similarly to the AUTOC in some cases as demonstrated in this study. However, since AUTOC is based on the first priority of treatment efficacy, it functions as the theoretical ‘gold standard’ or ruler for IQA in radiation therapy, by which other metrics can be evaluated for their applicability.

The optimization of CT simulation images for an external beam prostate treatment process has been used as an example to demonstrate the proposed IQA framework, but not the only use for this framework. We aim to re-introduce the IQA-in-RT theory through this preliminary study, which consists of developing the simulation framework itself, and then applying the framework to a simple case (CT simulation for prostate cancer patients), as a kind of feasibility study.

In this preliminary study, only the imaging and image segmentation steps were considered. The randomness in treatment plan generation and treatment response, patient motion during treatment, and internal organ motion can also be factored into the IQA-in-RT framework. As explained in Barrett *et al* (2013), equations (2) and (3) can be tailored to analyze varied segmentation algorithms and patient motion, as well as other components within the RT workflow that might affect TCP and NTCP.

As indicated in the IQA-in-RT theory, an important random effect that arises in task-based IQA is patient-to-patient variability. Our implemented IQA-in-RT framework addresses this issue by employing SOMs. Learning-based SOMs are generated to fully characterize the statistical structural variations of anatomical OARs and tumor targets within a specific ensemble of cancer patients. The SOMs can facilitate accurate modeling of patient anatomy

in computer-simulation studies that are conducted to optimize RT imaging systems and algorithms. In RT applications, databases of high-quality volumetric images and organ contours are available that can facilitate this SOM development. Our proposed methodology learns geometric attribute distributions (GAD) of human anatomic organs from a broad patient population. Therefore, the randomness in organ shape and position reflects the learned variability of human anatomy.

Many different imaging modalities are employed in radiation therapy for a variety of purposes. The proposed general framework has the potential to be applied to assess other imaging systems and processing algorithms designed for other images, such as MRI simulation images for treatment planning and on-board MRI or cone-beam CT images for patient positioning. After incorporation into the currently developed IQA-in-RT framework, any imaging modality could be assessed based on RT tasks. Specifically, image quality can be assessed for both currently challenging imaging modalities, such as CBCT (due to photon scatter), and newly developed technologies, such as MRI-guided radiation therapy. Additionally, by creating SOMs that consider patient motion, this framework has the potential to be employed to assess 4DCT or 4DMRI images for motion management, which is one of our on-going research works.

Figure 6 displays sample TCP, NTCP, and TOC curves calculated by the implemented simulation framework. In this study, the therapeutic ratio (TCP/NTCP for a given α value) is much improved over ratios in published works on clinical IMRT treatment plans (Boersma *et al* 1998, De Meerleer *et al* 2000, Witte *et al* 2007). Still, the current framework implementation presents a more idealized scenario than real clinical implementations. There are four main contributions to this decreased realism. First, since both patient motion and setup errors were not incorporated into this initial framework, no planning margin was added to the prostate. It is assumed in this study that the delivered dose is equal to the planned dose. Considering the impact of intra- and inter-fraction motion on the delivered dose will improve the realism of the proposed framework. For example, incorporation of a margin will certainly increase the dose to surrounding normal tissues and degrade treatment plan quality. Second, it was observed qualitatively that the image contrast between soft tissues (prostate, rectum, and bladder) was enhanced when compared to clinically acquired images; this enhancement is a direct result of the choice of attenuation coefficients for the CT projection simulation, as described in section 4.2.1 Third, no calibration to an actual linear accelerator was performed, and therefore no determination was made concerning the deliverability of the optimized treatment plan. Finally, it was also observed qualitatively that the prostate-bladder-rectum organ spacing was greater than observed in clinical images, in large part due to the method of organ intersection removal as described in the SOM creation (Dolly *et al* 2018). In future work, we plan to test many of these hypotheses and improve the framework.

In this framework, the TOC curve and AUTO C were used as surrogates for the final radiation treatment task. However, the TOC curve might be too sensitive to the dose offset. In addition, using an AUTO C with a single value might not capture information about the pattern of trade-offs of the particular discriminator algorithm. These issues will be further analyzed and investigated in future work.

Within the developed IQA-in-RT framework shown in figure 1, the individual modules described in section 3 can also be improved. First, to increase simulation automation, it is desirable to integrate automatic segmentation into the framework without use of third party applications. This could be accomplished using the aforementioned ITK library (Yoo *et al* 2002). Second, while the static IMRT treatment implemented in this study is common, patient motion during radiation treatment plan delivery can severely affect patient treatment. Motion management should be included to more accurately assess image quality. Finally, while the AUTO C is presented here as the FOM, there is some question as to the viability of the *in vitro* radiobiological models in comparison to *in vivo* patient results; other reported treatment outcome metrics (Chao *et al* 2003, Strigari *et al* 2008) will be assessed as viable FOMs in our future work.

Currently, the quality assurance (QA) methods of simulation CT in the RT workflow are performed with traditional IQA metrics, which have their advantages and benefits. However, these traditional metrics have certain limitations. In the field of diagnostic imaging, it is widely accepted that these traditional, physical IQ measures can lead to suboptimal designs of imaging systems because: (1) they do not account for all sources of randomness in the measurement data, and (2) they ignore the intended use (tasks) of the images. Future work for this project could potentially include routine quality assurance, for example establishing QA baselines/tolerances that are directly correlated to AUTO C.

7. Conclusion

In this work, a practical and modular framework was presented for implementing the IQA-in-RT theory for the assessment of imaging components on the basis of RT treatment outcomes. Computer-simulation studies were conducted to demonstrate the feasibility and utility of the proposed IQA-in-RT framework. This study provides a robust, fully-computer-simulated task-based IQA framework for radiation therapy. The presented work provides a practical solution for the further investigation and analysis of the IQA-in-RT theory and advances its applications in improving RT clinical practice and cancer patient care. The framework's flexibility allows for incorporating additional modules to include other aspects of the treatment process.

Acknowledgments

This work was supported in part by award NIH R01EB020604, R01EB023045, R01NS102213, and R21CA223799.

Appendix

A.1.

The NIST materials utilized for the patient object models

Material name	NIST material	ρ (g cm ⁻³)	Structures
Adipose	Adipose tissue (ICRU 44)	0.95	External
Bone (compact)	Bone, cortical (ICRU 44)	1.92	Femurs (exterior)
Bone (spongy)	Bone, cortical (ICRU 44)	1.0	Femurs (interior)

Material name	NIST material	ρ (g cm ⁻³)	Structures
Soft tissue	Tissue, soft (ICRU Four-Component)	1.0	Bladder, prostate, rectum

The NIST materials utilized for the patient object models are shown in table A1.

A.2.

EUD-based Radiobiological model parameters for calculating TCP and NTCP

Tissue	Tissue type	α_e	TD ₅₀ (Gy)	TCD ₅₀ (Gy)	γ_{50}
Prostate	Tumor	-10	—	28.34	1
Bladder	Normal	8.33	80	—	4
Rectum	Normal	2	80	—	4
Femur	Normal	4	65	—	4

The EUD-based radiobiological model parameters for calculating TCP and NTCP are shown in table A2.

A.3.

List of acronyms used in the paper

acronyms in the paper	
AUTOC	Area under therapeutic operating characteristic curve
CNR	Contrast to noise ratio
CT	Computed tomography
CTDI	Computed tomography dose index
DPM	Dose planning machine
DSC	Dice's similarity coefficient
DVH	Dose volume histogram
EUD	Equivalent uniform dose
FBP	Filtered back projection
FOM	Figure of merit
GAD	Geometric attribute distribution
IMRT	Intensity modulated radiation therapy
IQA	Image quality assessment
kVp	kilo Voltage peak (kVp)
MTF	Modulation transfer function
NEQ	Noise equivalent quanta
NPS	Noise power spectrum
NTCP	Normal tissue complication probability
OAR	Organs at risk
ROC	Receiver operating characteristic curve
ROI	Region of interest
RT	Radiation therapy
SOM	Stochastic object model
SSIM	Structure similarity index
TCP	Tumor control probability
TOC	Therapeutic operating characteristic curve

References

- Ågren A, Brahme A and Turesson I 1990 Optimization of uncomplicated control for head and neck tumors *Int. J. Radiat. Oncol. Biol. Phys* 19 1077–85
- Andrews JR 1985 Benefit, risk, and optimization by ROC analysis in cancer radiotherapy *Int. J. Radiat. Oncol. Biol. Phys* 11 1557–62
- Barrett HH and Myers KJ 2013 *Foundations of Image Science* (New York: Wiley)
- Barrett HH, Denny J, Wagner RF and Myers KJ 1995 Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance *J. Opt. Soc. Am. A* 12 834–52
- Barrett HH, Kupinski MA, Müller S, Halpern HJ, Morris JC III and Dwyer R 2013 Objective assessment of image quality VI: imaging in radiation therapy *Phys. Med. Biol* 58 8197 [PubMed: 24200954]
- Barrett HH, Myers KJ, Hoeschen C, Kupinski MA and Little MP 2015 Task-based measures of image quality and their relation to radiation dose and patient risk *Phys. Med. Biol* 60 R1
- Barrett HH, Wilson DW, Kupinski MA, Aguwa K, Ewell L, Hunter R and Müller S 2010 Therapy operating characteristic (TOC) curves and their application to the evaluation of segmentation algorithms *Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment* vol 7627 (International Society for Optics and Photonics)
- Boersma LJ, Van den Brink M, Bruce AM, Shouman T, Gras L, te Velde A and Lebesque JV 1998 Estimation of the incidence of late bladder and rectum complications after high-dose (70–8 Gy) conformal radiotherapy for prostate cancer, using dose-volume histograms *Int. J. Radiat. Oncol. Biol. Phys* 41 83–92 [PubMed: 9588921]
- Boone JM and Seibert JA 1997 An accurate method for computer-generating tungsten anode x-ray spectra from 30 to 140 kv *Med. Phys* 24 1661–70 [PubMed: 9394272]
- Caselles V, Kimmel R and Sapiro G 1997 Geodesic active contours *Int. J. Comput. Vis* 22 61–79
- Chao KC, Blanco AI and Dempsey JF 2003 A conceptual model integrating spatial information to assess target volume coverage for IMRT treatment planning *Int. J. Radiat. Oncol. Biol. Phys* 56 1438–49 [PubMed: 12873690]
- Dawson LA and Jaffray DA 2007 Advances in image-guided radiation therapy *J. Clin. Oncol* 25 938–46 [PubMed: 17350942]
- De Meerleer GO, Vakaet LA, De Gerssem WR, De Wagter C, De Naeyer B and De Neve W 2000 Radiotherapy of prostate cancer with or without intensity modulated beams: a planning comparison *Int. J. Radiat. Oncol. Biol. Phys* 47 639–48 [PubMed: 10837946]
- Deasy JO, Mayo CS and Orton CG 2015 Treatment planning evaluation and optimization should be biologically and not dose/volume based *Med. Phys* 42 2753 [PubMed: 26127027]
- Dolly S, Chen HC, Anastasio M, Mutic S and Li H 2016 Practical considerations for noise power spectra estimation for clinical CT scanners *J. Appl. Clin. Med. Phys* 17 392–407
- Dolly SR, Lou Y, Anastasio MA and Li H 2018 Learning-based stochastic object models for characterizing anatomical variations *Phys. Med. Biol* 63 065004 [PubMed: 29536945]
- Drzymala R, Mohan R, Brewster L, Chu J, Goitein M, Harms W and Urie M 1991 Dose-volume histograms *Int. J. Radiat. Oncol. Biol. Phys* 21 71–8 (Three-dimensional photon treatment planning report of the collaborative working group on the evaluation of treatment planning for external photon beam radiotherapy) [PubMed: 2032898]
- Emami B, Lyman J, Brown A, Cola L, Goitein M, Munzenrider J, Shank B, Solin L and Wesson M 1991 Tolerance of normal tissue to therapeutic irradiation *Int. J. Radiat. Oncol. Biol. Phys* 21 109–22 [PubMed: 2032882]
- Friedman SN, Fung GS, Siewerdsen JH and Tsui BM 2013 A simple approach to measure computed tomography (CT) modulation transfer function (MTF) and noise-power spectrum (NPS) using the american college of radiology (ACR) accreditation phantom *Med. Phys* 40 051907 [PubMed: 23635277]
- Frigo M and Johnson SG 1998 FFTW: an adaptive software architecture for the FFT *Proc. of the 1998 IEEE Int. Conf. on Acoustics, Speech and Signal Processing* vol 3 (IEEE) pp 1381–4

- Gabriel E et al. 2004 Open MPI: goals, concept, and design of a next generation MPI implementation European Parallel Virtual Machine/Message Passing Interface Users' Group Meeting (Springer) pp 91–104
- Gay HA and Niemierko A 2007 A free program for calculating EUD-based NTCP and TCP in external beam radiotherapy Phys. Med 23 115–25 [PubMed: 17825595]
- Hall EJ and Wu CS 2003 Radiation-induced second cancers: the impact of 3d-CRT and IMRT Int. J. Radiat. Oncol. Biol. Phys 56 83–8 [PubMed: 12694826]
- Hanley JA and McNeil BJ 1982 The meaning and use of the area under a receiver operating characteristic (roc) curve Radiology 143 29–36 [PubMed: 7063747]
- Hoffmann AL, Huizenga H and Kaanders JH 2013 Employing the therapeutic operating characteristic (TOC) graph for individualised dose prescription Radiat. Oncol 8 55 [PubMed: 23497640]
- Ibanez L, Schroeder W, Ng L and Cates J 2005 The ITK software guide: updated for ITK version 2.4 (New York: Kitware)
- Jaffray DA GM 2015 Radiation therapy for cancer Cancer: Disease Control Priorities vol 3, 3rd edn (Washington, DC: The World Bank)
- Kak AC and Slaney M 2001 Principles of Computerized Tomographic Imaging (Philadelphia, PA: SIAM)
- Khan FM and Gerbi BJ 2012 Treatment Planning in Radiation Oncology (Dordrecht: Kluwer)
- Li H, Yu L, Anastasio MA, Chen HC, Tan J, Gay H, Michalski JM, Low DA and Mutic S 2014 Automatic CT simulation optimization for radiation therapy: a general strategy Med. Phys 41 031913 [PubMed: 24593731]
- Lyman JT 1985 Complication probability as assessed from dose-volume histograms Radiat. Res 104 S13–9
- Metz CE, Tokars RP, Kronman HB and Griem ML 1982 Maximum likelihood estimation of dose-response parameters for therapeutic operating characteristic (TOC) analysis of carcinoma of the nasopharynx Int. J. Radiat. Oncol. Biol. Phys 8 1185–92 [PubMed: 7118618]
- Möller T and Trumbore B 2005 Fast, minimum storage ray/triangle intersection ACM SIGGRAPH 2005 Courses (ACM)
- Moore DH and Mendelsohn ML 1972 Optimal treatment levels in cancer therapy Cancer 30 97–106 [PubMed: 5040749]
- Niemierko A 1997 Reporting and analyzing dose distributions: a concept of equivalent uniform dose Med. Phys 24 103–10 [PubMed: 9029544]
- Njeh CF, Parker BC and Orton CG 2015 Evaluation of treatment plans using target and normal tissue dvhs is no longer appropriate Med. Phys 42 2099–102 [PubMed: 25979004]
- Okunieff P, Morgan D, Niemierko A and Suit HD 1995 Radiation dose-response of human tumors Int. J. Radiat. Oncol. Biol. Phys 32 1227–37 [PubMed: 7607946]
- Otto K 2008 Volumetric modulated arc therapy: IMRT in a single gantry arc Med. Phys 35 310–7 [PubMed: 18293586]
- Rana S and Cheng C 2014 Radiobiological impact of planning techniques for prostate cancer in terms of tumor control probability and normal tissue complication probability Ann. Med. Health Sci. Res 4 167–72 [PubMed: 24761232]
- Rice JA 2006 Mathematical Statistics and Data Analysis (Boston, MA: Cengage Learning)
- Sempau J, Wilderman SJ and Bielajew AF 2000 Dpm, a fast, accurate monte carlo code optimized for photon and electron radiotherapy treatment planning dose calculations Phys. Med. Biol 45 2263 [PubMed: 10958194]
- Shepard D, Earl M, Li X, Naqvi S and Yu C 2002 Direct aperture optimization: a turnkey solution for step-and-shoot IMRT Med. Phys 29 1007–18 [PubMed: 12094970]
- Siegel RL, Miller KD and Jemal A 2015 Cancer statistics CA: Cancer J. Clinicians 65 5–29
- Strigari L, D'Andrea M, Abate A and Benassi M 2008 A heterogeneous dose distribution in simultaneous integrated boost: the role of the clonogenic cell density on the tumor control probability Phys. Med. Biol 53 5257 [PubMed: 18758004]
- Taguchi K and Aradate H 1998 Algorithm for image reconstruction in multi-slice helical CT Med. Phys 25 550–61 [PubMed: 9571623]

- Tokars RP and Griem ML 1979 Carcinoma of the nasopharynx an optimization of radiotherapeutic management for tumor control and spinal cord injury *Int. J. Radiat. Oncol. Biol. Phys* 5 1741–8
- US Department of Commerce N 2009 NIST: x-ray mass attenuation coefficients (www.nist.gov/pml/data/xraycoef/index.cfm) (Accessed: 1 April 2016)
- Vanderstraeten B. et al. 2007; Conversion of CT numbers into tissue parameters for monte carlo dose calculations: a multi-centre study. *Phys. Med. Biol.* 52:539. [PubMed: 17228104]
- Wang Z, Bovik AC, Sheikh HR and Simoncelli EP 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process* 13 600–12 [PubMed: 15376593]
- Warkentin BJ, Stavrev P, Stavreva N, Field C and Fallone BG 2004 A TCP-NTCP estimation module using dvhs and known radiobiological models and parameter sets *J. Appl. Clin. Med. Phys* 5 50–63 [PubMed: 15753933]
- Webb S. 1991; Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator. *Phys. Med. Biol.* 36:1201. [PubMed: 1946603]
- Webb S and Nahum AE 1993 A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density *Phys. Med. Biol* 38 1923 [PubMed: 8108492]
- Witte MG, van der Geer J, Schneider C, Lebesque JV, Alber M and van Herk M 2007 IMRT optimization including random and systematic geometric errors based on the expectation of TCP and NTCP *Med. Phys* 34 3544–55 [PubMed: 17926957]
- Yan D, Vicini F, Wong J and Martinez A 1997 Adaptive radiation therapy *Phys. Med. Biol* 42 123 [PubMed: 9015813]
- Yoo T, Ackerman M, Lorensen W, Schroeder W, Chalana V, Aylward S, Metaxas D and Whitaker R 2002 Engineering and algorithm design for an image processing API: a Technical Report on ITK —The Insight Toolkit *Stud. Health Technol. Inform* 85 586–92 [PubMed: 15458157]
- Yushkevich PA, Piven J, Cody Hazlett H, Gimpel Smith R, Ho S, Gee JC and Gerig G 2006 User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability *NeuroImage* 31 1116–28 [PubMed: 16545965]

Implementation of the IQA-in-RT Theory

Figure 1.

The general framework of the IQA-in-RT theory implementation.

Figure 2.
The key steps in the stochastic object model (SOM) creation process.

Figure 3.

(a) One example realization of the pelvis SOM, including the prostate, bladder, rectum, femoral heads, pelvic bone, and seminal vesicles. The patient's external surface is not shown. (b) A 2D cross-section located at a cutplane in the 3D model.

Figure 4.

Examples of CT images generated using a 140 kVp source spectrum for varying photon numbers, corresponding to the realizations of the SOM shown in figure 3. The reconstruction filter was set as a *sinc* function with $\alpha = 2$, reconstruction field of view (FOV) = 45 cm, slice thickness = 3 mm, and helical pitch = 1. Panels (a)–(c) display the decrease in noise as the number of photons increases. (a) $I_0 = 5 \times 10^3$. (b) $I_0 = 5 \times 10^4$. (c) $I_0 = 5 \times 10^5$.

Figure 5. Optimized dose distribution on a 2D cross-section image (left) and a plot of the corresponding dose-volume histograms (right).

Figure 6.

The TCP, NTCP, and TOC curves for prostate tumor control versus rectal bleeding, a normal tissue complication of the treatment plan. The AUTOC value is 0.9929. (a) TCP/NTCP curves versus dose parameter λ . (b) TOC curve.

Figure 7.

The AUTOC is a function of the CT imaging dose, in terms of a relative CTDI_{vol}. The full curve is provided (left) along with a zoomed-in version (right) to help distinguish the higher AUTOC values. (a) Regular. (b) Zoomed.

Figure 8.

Simulated CT images with varied reconstruction filters. CT images were generated using a 140 kVp source spectrum and with the number of photons per detector element set to 5×10^3 , which corresponds to a relative $\text{CTDI}_{vol} = 3.05$. CT images were simulated corresponding to the realizations of the SOM shown in figure 3. The reconstructed field of view (FOV) = 45 cm, slice thickness = 3 mm, and helical pitch = 1. (a) $\alpha \approx 0$. (b) $\alpha = 4$. (c) $\alpha = 10$.

Figure 9. Results of using AUTOC as a FOM for optimizing CT image reconstructions by optimizing the filter parameter α . (a) Full curve. (b) Zoomed curve.

Figure 10.

The AUTOC compared to various IQA metrics for optimizing imaging dose and image reconstructions (the filter parameter a). (a) Optimizing CT imaging dose. (b) Optimizing FBP reconstruction filter.

Figure 11.

AUTOC versus radiobiological model analysis of bias within the AUTOC calculation. The upper two figures display the AUTOC calculations for different radiobiological models, while the lower two display the AUTOC calculated using different model parameters according to different prostate cancer stages. (a) Imaging dose. (b) Reconstruction filter. (c) Imaging dose. (d) Reconstruction filter.

Table 1.

Treatment prescription and DVH constraints for IMRT prostate treatment plan optimization.

Target/OARs	Constraint 1	Constraint 2
Prostate (tumor target)	Min. Dose 78 Gy	Max. Dose < 81 Gy
Bladder	V(70 Gy) < 30%	V(55 Gy) < 50%
Femur (L and R)	V(50 Gy) < 5%	—
Rectum	V(70 Gy) < 20%	V(50 Gy) < 50%