Objective assessment of image quality VI: imaging in radiation therapy

This content has been downloaded from IOPscience. Please scroll down to see the full text.
(http://iopscience.iop.org/0031-9155/58/22/8197)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 54.191.40.80
This content was downloaded on 11/06/2017 at 11:54

Please note that terms and conditions apply.

You may also be interested in:

Task-based measures of image quality and their relation to radiation dose and patient risk
Harrison H Barrett, Kyle J Myers, Christoph Hoeschen et al.

Dose response relations and treatment plans
Panayiotis Mavroidis, Bengt K Lind and Anders Brahme

Estimation of parameters of dose--volume models and their confidence limits
P van Luijk, T C Delvigne, C Schilstra et al.

Evaluation of dose-response models and parameters predicting radiation induced pneumonitis
Ioannis Tsougos, Panayiotis Mavroidis, Juha Rajala et al.

Monte Carlo role in radiobiological modelling of radiotherapy outcomes
Issam El Naqa, Piotr Pater and Jan Seuntjens

EUCLID: an outcome analysis tool for high-dimensional clinical studies
Olivier Gayou, David S Parda and Moyed Miften

Collimator optimization in myocardial perfusion SPECT using the ideal observer and realistic background variability for lesion detection and joint detection and localization tasks
Michael Ghaly, Yong Du, Jonathan M Links et al.

Nuclear physics in particle therapy: a review
Marco Durante and Harald Paganetti

A radiobiological analysis of the effect of 3D versus 4D image-based planning
Teboh Roland, Panayiotis Mavroidis, Alonso Gutierrez et al.
Objective assessment of image quality VI: imaging in radiation therapy

Harrison H Barrett\textsuperscript{1,2}, Matthew A Kupinski\textsuperscript{1,2}, Stefan Müller\textsuperscript{3}, Howard J Halpern\textsuperscript{4}, John C Morris III\textsuperscript{5} and Roisin Dwyer\textsuperscript{6}

\textsuperscript{1} College of Optical Sciences, University of Arizona, Tucson, AZ, USA
\textsuperscript{2} Center for Gamma-Ray Imaging, Department of Medical Imaging, University of Arizona, Tucson, AZ, USA
\textsuperscript{3} University of Duisberg-Essen, Essen, Germany
\textsuperscript{4} Center for EPR Imaging In Vivo Physiology, Department of Radiation and Cellular Oncology, University of Chicago, Chicago, IL, USA
\textsuperscript{5} Department of Endocrinology, Mayo Clinic, Rochester, MN, USA
\textsuperscript{6} Discipline of Surgery, School of Medicine, National University of Ireland Galway, Ireland

E-mail: barrett@radiology.arizona.edu

Received 18 June 2013, in final form 18 August 2013
Published 7 November 2013
Online at stacks.iop.org/PMB/58/8197

Abstract

Earlier work on objective assessment of image quality (OAIQ) focused largely on estimation or classification tasks in which the desired outcome of imaging is accurate diagnosis. This paper develops a general framework for assessing imaging quality on the basis of therapeutic outcomes rather than diagnostic performance. By analogy to receiver operating characteristic (ROC) curves and their variants as used in diagnostic OAIQ, the method proposed here utilizes the therapy operating characteristic or TOC curves, which are plots of the probability of tumor control versus the probability of normal-tissue complications as the overall dose level of a radiotherapy treatment is varied. The proposed figure of merit is the area under the TOC curve, denoted AUTOC. This paper reviews an earlier exposition of the theory of TOC and AUTOC, which was specific to the assessment of image-segmentation algorithms, and extends it to other applications of imaging in external-beam radiation treatment as well as in treatment with internal radioactive sources. For each application, a methodology for computing the TOC is presented. A key difference between ROC and TOC is that the latter can be defined for a single patient rather than a population of patients.

(Some figures may appear in colour only in the online journal)
1. Introduction

Imaging is an integral part of the planning and execution of radiation therapy. Images are essential in the initial diagnosis and localization of a neoplasm; in planning a course of radiation therapy; in correcting for patient motion and inaccurate patient positioning during the therapy, and in monitoring the response of the tumor to the therapy. Successful therapeutic outcomes therefore depend critically on the quality of the images and image-analysis methods.

The paradigm of objective assessment of image quality (OAIQ) is based on the premise that image quality should be defined by the ability of a user to perform medically or scientifically relevant tasks with the image data. The theory underlying OAIQ is detailed in earlier papers in this series (Barrett 1990, Barrett et al 1995, 1998, 2006, Caucci and Barrett 2012) and in Barrett and Myers (2004). For diagnostic tasks, many evaluation methods related to receiver operating characteristic (ROC) curves have been developed and used to assess imaging systems and reconstruction algorithms (Chakraborty and Berbaum 2004, Clarkson 2007). Much less has been done with the estimation tasks that arise in the context of radiation therapy, and it is the purpose of the present paper to fill this gap.

The overall goal of radiation therapy is to maximize the probability of destroying tumors while minimizing damage to surrounding normal tissues. Many sophisticated approaches to this goal, including three-dimensional conformal radiotherapy (Zelevsky et al 1998); intensity-modulated radiotherapy (Eisbruch 2002, Hatano et al 2007, Ahnesjo et al 2006), image-guided radiotherapy (Jaffray 2006), brachytherapy and radioimmunotherapy, have been developed for this purpose, but they all depend on information extracted from images of the patient. Moreover, even after the treatment plan is finalized, there is still a tradeoff between the probability of tumor control and the probability of normal-tissue complications, both of which increase when the overall dose of the radiation is increased. In external-beam radiotherapy, this overall dose is controlled by the beam current or the duration of each treatment fraction, and in brachytherapy or radioimmunotherapy, the overall dose is proportional to the activity of the internal radioactive source.

In the radiotherapy literature, tumor-control probability is referred to as TCP and normal-tissue complication probability is called NTCP, but we prefer the more mathematical notations, Pr(TC) and Pr(NTC), respectively, where Pr(·) denotes a probability, as distinct from a probability density function (PDF), which we denote as pr(·).

Much work has gone into estimating Pr(TC) and Pr(NTC) (Chow et al 2008, Romeijn and Dempsey 2008, Wu et al 2002, Stavrev et al 2009, Deasy and El Naqa 2008, Karger 2006, Kutcher 1996, Park et al 2005, Baumann and Petersen 2005, Stavrev et al 2001, Schultheiss et al 1983, Niemierko and Goitein 1993, Lyman and Wolbarst 1987, 1989), and standardized modules for their calculation are freely available (Warkentin et al 2004, Gay and Niemierko 2007, Sanchez-Nieto and Nahum 2000). The module developed by Warkentin et al (2004), for example, uses a planned dose distribution characterized by a dose-volume histogram (DVH) to calculate the probabilities. For Pr(NTC), it assumes that organs are composed of functional subunits, and that organ function is compromised when a certain critical fraction of the subunits is damaged by radiation. These assumptions lead to a cumulative binomial dose-response curve with two free parameters. Similarly, Pr(TC) in the Warkentin module is computed with either a Poisson model or a radiobiological model that incorporates parameters describing linear-quadratic cell kill and repopulation. The module includes a database of parameter values for different normal tissues and tumor types as well as for different degrees of complication.

A similar module, developed by Gay and Niemierko (2007), uses a dose distribution described by the equivalent uniform dose rather than DVH, and it bases its estimates of Pr(NTC) on published data for the radiation tolerance of normal tissues. As in Warkentin et al (2004), different parameters are given for different normal tissues and different degrees
of complication. For ear damage, for example, different parameter values are given for acute serous otitis and chronic serous otitis.

The form of the output from such modules is depicted schematically in figure 1. The graphs on the left are for a hypothetical therapy plan where the operating point (overall dose) can be chosen for high probability of tumor control and minimal probability of some specified normal-tissue complication. In the graphs on the right, however, \( \text{Pr}(\text{NTC}) \) is higher for any given choice of \( \text{Pr}(\text{TC}) \) than in the graphs on the left, so these curves correspond to a less effective treatment plan, at least for this one particular kind and degree of complication.

The curves in figure 1 are directly analogous to plots of true-positive fraction (probability of detection) and false-positive fraction (probability of false alarm) versus a decision threshold in signal-detection problems (Barrett and Myers 2004). It is an easy step to replot figure 1 as \( \text{Pr}(\text{TC}) \) versus \( \text{Pr}(\text{NTC}) \) with the overall dose as a hidden parameter. This plot, initially introduced by Moore and Mendelsohn (1972) and analyzed in ROC terms by Metz et al (1982), is referred to as the therapy operating characteristic or TOC curve. The TOC curves corresponding to figure 1 are shown in figure 2. TOC curves were further studied by Andrews (1985) who discussed how the TOC concept could be used for objective therapy optimization on a per-patient basis.

This paper uses the area under the TOC curve (AUTOC) as a figure of merit for radiation therapy planning and for the imaging components that affect the plan. An AUTOC of 1.0 is ideal, and the diagonal line with AUTOC = 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. In contrast to ROC, an AUTOC < 0.5 can also occur, in the case of a very bad treatment plan where there is a large probability of damage to normal tissues with little chance of tumor control.

In section 2 we review the theory of TOC as originally laid out in Barrett et al (2010). In section 3 we discuss some simplifying assumptions applicable to external-beam radiation therapy. In section 4 we show how AUTOC can be used to evaluate segmentation algorithms, and in section 5, we show how it can be used to evaluate the effect of patient motion on external-beam radiation therapy. Application of the theory to the case where molecular imaging (MI) data are included in the planning of external-beam radiotherapy is discussed in section 6, and application to therapy with targeted internal radiation sources is illustrated in section 7 by discussing therapy based on the human sodium iodine symporter.
Figure 2. Therapy operating characteristic curves for the two treatment plans depicted in figure 1. The upper curve in this figure corresponds to the plan on the left in figure 1.

In all of these applications, the theory is couched in the language of conditional probabilities and formally stated as high-dimensional nested integrals, starting with a fully general expression and then making reasonable assumptions to get tractable results. The assumptions are often stated conceptually in the form of high-dimensional delta functions, which should be regarded as devices for showing that certain integrals are unnecessary if the stated assumptions can be justified. In practice, all remaining integrals can be performed by Monte Carlo methods, and none of the integrals will have to be performed analytically. We emphasize that it is not our intent to evaluate the effects of all possible sources of uncertainty in the radiation therapy process simultaneously, nor do we imply that the TOC is a complete measure of the efficacy of the therapy. Rather, in each of the sections that follow, we focus on one aspect of imaging in radiation therapy at a time and consider mainly the variability arising from the images, since they are the concern of this paper. If we have two or more competing approaches to the imaging task in question (e.g., different segmentation algorithms), this procedure will allow us to rank them in terms of AUTOC. We return to this point in section 8 where we discuss how other sources of randomness might change this rank ordering.

For a worked example of how this formalism can be applied to the evaluation of segmentation methods, see Barrett et al (2010).

2. General theory

A key feature that sets tasks relevant to radiation therapy apart from other tasks used in OAIQ is the importance of boundaries. Given an image data set, the first step in the planning of external-beam therapy is to segment the images and estimate the boundaries of the tumor and of any nearby normal tissues that might be subject to damage from the therapy radiation. Then the treatment plan strives to maximize the radiation dose within the tumor boundary while minimizing the dose within the boundaries of sensitive normal tissues. When the planned
therapy is administered, it is also important to assure that the boundaries of the tumor remain within the planned treatment volume. Therefore the main goal of this section is to determine the relation between the boundaries and the probabilities of tumor control and normal-tissue complications.

The imaging system that provides the data used in boundary determination is modeled as a continuous-to-discrete mapping (Barrett and Myers 2004) where the object is a function of continuous variables and the image is a discrete set of numbers such as gray-level values at voxels. We assume that the tumor and all normal organs of interest have well-defined boundaries which can be represented as mathematical surfaces in a 3D space. For patient $j$, these surfaces are specified by a (potentially infinite) set of parameters denoted as $B_j$.

The surface descriptions contained in $B_j$ are, of course, unknown, and it is the goal of a segmentation algorithm to estimate them in some way. The resulting estimate of the surfaces for patient $j$ is denoted $\hat{B}_j$. Unlike the continuous descriptions in $B_j$, the estimates in $\hat{B}_j$ specify a set of voxels or surface tesselations that define the surfaces in a discrete sense.

The data used for estimating the boundaries consist of one or more images. Typically, the normal-organ boundaries and the tumor boundaries will be estimated from a CT or MR image, but there is an increasing interest in including biological information estimated from PET images into the treatment planning process. For generality, we denote the available image data set for patient $j$ as $G_j$, but we assume that this set must always include a CT study because x-ray attenuation data are needed for treatment planning.

From the estimated boundaries and the original CT data, a dosimetrist or physicist will derive a treatment plan which specifies the beam profiles, energies and exposures. Implementation of this plan will result in a dose distribution for patient $j$ denoted $D_j(r)$, where $r$ is a three-dimensional position vector in the patient’s body. We assume that the dose distribution scales by the same factor at each position $r$ as the exposure time or beam current is varied. If we assume further, for fractionated treatments, that the exposure is varied by the same proportion in each session, we can write

$$D_j(r) = \alpha D_{0j}(r),$$

where $D_{0j}(r)$ is the dose distribution at some reference exposure and $\alpha$ is the ratio of an actual exposure to the reference exposure. Equivalently, we can say that $D_{0j}(r)$ is the dose distribution for $\alpha = 1$. For notational economy, we denote $\alpha D_{0j}(r)$ as $D_j$, with the dependence on $\alpha$ implicit. One can think of $\alpha$ as setting the operating point on the TOC curve, hence setting the tradeoff between probability of tumor control and probability of occurrence of some specified normal tissue complication. It is important to note that the entire TOC curve is generated by varying $\alpha$, so the shape of the TOC curve and the area under it are independent of $\alpha$, though of course they depend on the spatial distribution of the dose. Thus $\alpha$ controls the abscissa in figure 1 but disappears in figure 2; instead any particular choice of the scale factor $\alpha$ determines one point on the TOC curve.

The stochastic quantities that control the probabilities of tumor control and normal-tissue complications are noise in the image data; randomness in the estimated boundaries derived from a given image set; uncertainty in the actual delivered dose distribution, and of course the stochastic nature of the tissue response itself. We can describe all of these effects abstractly in terms of conditional PDFs.

The multivariate PDF for the image noise for patient $j$, denoted $\text{pr}(G_{j}|j)$, describes the effect of the noise in the raw projection data as propagated through the reconstruction algorithm. The conditioning on $j$ implies that $\text{pr}(G_{j}|j)$ describes the stochastic variation of the images for hypothetical repeated imaging of the same patient with all imaging parameters (including the radiotracer distribution in the case of PET or SPECT) held constant; for much
The PDF for the estimated boundaries given the image data, denoted $\Pr(\hat{B}_j|G_j)$, would be a huge-dimensional delta function for an automatic segmentation algorithm because repeated trials with the same images would always deliver the same boundaries, but human analysts would not be so repeatable. We therefore retain $\Pr(\hat{B}_j|G_j)$ as an unspecified PDF for now.

The PDF for the actual dose for patient $j$ given the estimated boundaries and the image data is denoted $\Pr(D_j|\hat{B}_j, G_j, j)$. The estimated boundaries $\hat{B}_j$ and the variations in x-ray attenuation coefficient in the CT data $G_j$ determine the treatment plan and hence the intended dose distribution, but this PDF allows the actual dose to deviate in an unknown, random way from the intended one. The remaining conditioning on $j$ allows for the response to depend on characteristics of the patient not captured in the image data and hence not considered in the calculations of probabilities of tumor control and normal-tissue complications.

Finally, the probability (not PDF) of tumor control, which depends on the actual dose distribution, the actual boundaries and possibly other biological factors such as oxygenation and perfusion, can be denoted $\Pr(TC|D_j, B_j, j)$. The notation here is a trifle redundant because conditioning on the patient is implicitly specifying the true organ boundaries, but we retain the dependence on $B_j$ to emphasize its importance.

With these definitions, we can apply elementary rules of conditional probability to write the probability of tumor control for patient $j$ as

$$\Pr(TC|\alpha, j) = \int dD_j \int d\hat{B}_j \int dG_j \Pr(TC|D_j, \hat{B}_j, G_j, j) \Pr(D_j|\hat{B}_j, G_j, j) \Pr(\hat{B}_j|G_j) \Pr(G_j|j).$$

(2)

where the dependence on $\alpha$ comes from (1).

Similarly, the general expression for the probability of normal-tissue complications is

$$\Pr(NTC|\alpha, j) = \int dD_j \int d\hat{B}_j \int dG_j \Pr(NTC|D_j, \hat{B}_j, G_j, j) \Pr(D_j|\hat{B}_j, G_j, j) \Pr(\hat{B}_j|G_j) \Pr(G_j|j).$$

(3)

Note that the probabilities given by (2) and (3) are determined by both the actual and the estimated boundaries; the estimated boundaries determine the treatment plan, but given a plan and hence a dose distribution, it is the actual boundaries, along with the overall dose $\alpha$, that determine $\Pr(TC|\alpha, j)$ and $\Pr(NTC|\alpha, j)$.

3. Simplifications relevant to external-beam radiation therapy

The general equations (2) and (3) involve complicated, high-dimensional integrals, but fortunately we can make some simplifying assumptions. Consider first the conditional PDF $\Pr(D_j|\hat{B}_j, G_j, j)$ which appears in (2) and (3). If we assume merely that the dose-calculation software is non-random, i.e., that it always returns the same computed dose distribution when it is rerun with the same estimated boundaries and image data, then this PDF must be a high-dimensional delta function.

Moreover, it is reasonable to neglect the potential inaccuracies in dose calculations and assume that the actual dose distribution, conditional on the estimated boundaries and the image data that go into the plan, is to a good approximation what the therapy planning program says it is. There are two main justifications for this assumption. First, this critical aspect of radiotherapy has been widely studied, and the dose-computation algorithms have been continually refined and validated by Monte Carlo studies (Chetty et al 2007). One can surely have confidence in them if an accurate image model is used. Second, the dose delivered
by a beam of high-energy photons is rather non-local; a photon that undergoes a photoelectric
or Compton interaction at a specific point in the tissue can deliver its dose over a range of a
centimeter or more, and there are many such interactions, so noise or limited spatial resolution
in the image may have little significant effect on the dose distribution.

The importance of this assumption is that we can eliminate one of the three integrals in
(2) or (3). If there are no significant inaccuracies in dose calculation itself, conditional on the
estimated boundaries and the image data, we can write

$$ \text{pr}(D_j | \hat{B}_j, G_j, j) = \delta[D_j - D(\hat{B}_j, G_j)]. $$

where $D(\hat{B}_j, G_j)$ is the dose distribution (interpreted as a function of continuous spatial
variables) returned by the planning algorithm. The delta function can then be used to perform
the integral over $D_j$, and (2) becomes

$$ \text{Pr}(TC|\alpha, j) = \int d\hat{B}_j \int dG_j \text{Pr}(TC|D(\hat{B}_j, G_j), B_j, j) \text{pr}(\hat{B}_j|G_j)\text{pr}(G_j|j). $$

It is important to note that the probability of tumor control still depends on the actual patient
boundaries $B$ in this integral; our assumption of accurate dose calculation is conditional on
the estimated boundaries, and it does not rule out additional dose-delivery errors from patient
motion, incorrect patient positioning or inaccurate segmentation.

Another reasonable approximation is to neglect the patient-to-patient variation of tissue
damage by ionizing radiation. As noted above, $\text{Pr}(TC|D_j, B_j, j)$ can depend on patient
characteristics not captured by the image data; that was the reason for the final conditioning
on $j$. In fact, we have little way of knowing what these characteristics are in clinical practice,
and we certainly could not incorporate them realistically in a simulation study. Thus we opt
to delete the last $j$ and write $\text{Pr}(TC|D_j, B_j) = \text{Pr}(TC|D_j, B_j)$.

Now we note that the product of the last two conditional probabilities in (5) can be
rewritten as a joint probability:

$$ \text{pr}(\hat{B}_j|G_j)\text{pr}(G_j|j) = \text{pr}(\hat{B}_j, G_j|j). $$

Thus (5) becomes

$$ \text{Pr}(TC|\alpha, j) = \langle \text{Pr}(TC|D(\hat{B}_j, G_j), B_j, j) \rangle |_{\hat{B}_j, G_j}. $$

where the notation indicates that the quantity in angle brackets is to be averaged over the joint
distribution of the subscripted quantities, $B_j$ and $G_j$.

4. Application to the evaluation of segmentation methods

Image segmentation can be performed manually by a human operator tracing boundaries on
an image, or by an automated segmentation algorithm, or by an automated algorithm where
the human intervenes to correct errors. These three options have analogies in the problem of
lesion detection from images, where the detection can be performed entirely by a human, by
a computer detection algorithm, or by a human in conjunction with an algorithm.

For the detection task, image quality is assessed by using a set of images, some of which
contain one or more lesions and some contain no lesion. It is important to the evaluation
that the true lesion status of each image be known to the experimenter, though of course
this information is not given to the observer. The are two common methods for generating
images with known truth; either the images can be completely simulated, or normal clinical
images can be used with simulated lesions inserted into them. As simulation models get ever
more realistic, the first option becomes more frequently the norm for lesion detection, and we
assume here that image-segmentation methods will also be assessed from realistic simulated
images. For a discussion of the evaluation of segmentation algorithms with clinical image and an example related to radiation therapy of the prostate, see Barrett et al (2010).

The expression in (7) suggests a Monte Carlo algorithm that can be readily implemented with simulated data. Suppose we begin with an anatomical phantom to represent a particular patient \( j \) and we generate \( K \) random image samples, differing only by the realization of the Poisson noise in the raw projection data. The \( k \)-th such image sample is denoted \( \mathbf{G}_{j}^{(k)} \) and the corresponding boundary estimate is denoted \( \hat{\mathbf{B}}_{j}^{(k)} \). The pair \((\hat{\mathbf{B}}_{j}^{(k)}, \mathbf{G}_{j}^{(k)})\) is a sample from the joint distribution \( \text{pr}(\hat{\mathbf{B}}_{j}, \mathbf{G}_{j}|j) \); only one index \( k \) is needed to specify the sample because \( \hat{\mathbf{B}}_{j}^{(k)} \) is generated directly (though not necessarily deterministically) from \( \mathbf{G}_{j}^{(k)} \).

Because we started with an anatomical phantom, the true boundaries \( \mathbf{B}_{j} \) are known, and we can use a standard therapy planning algorithm and standard TCP and NTCP modules to evaluate the quantity in angle brackets in (7). The Monte Carlo estimate of \( \text{Pr}(TC|\alpha, j) \) is then

\[
\hat{\text{Pr}}(TC|\alpha, j) = \frac{1}{K} \sum_{k=1}^{K} \text{Pr}[TC|D(\hat{\mathbf{B}}_{j}^{(k)}, \mathbf{G}_{j}^{(k)}), \mathbf{B}_{j}],
\]

where the probability on the right-hand side is interpreted as the output of the TCP module. Similarly,

\[
\hat{\text{Pr}}(NTC|\alpha, j) = \frac{1}{K} \sum_{k=1}^{K} \text{Pr}[NTC|D(\hat{\mathbf{B}}_{j}^{(k)}, \mathbf{G}_{j}^{(k)}), \mathbf{B}_{j}].
\]

The Monte Carlo estimates of (8) and (9) are still functions of \( \alpha \), so the computation gives one point on the TOC curve. Repeating for several values of \( \alpha \) as in an ROC study generates the TOC curve for phantom \( j \); the study can be repeated for several phantoms to get an average TOC curve and AUTOC.

Evaluation of automatic segmentation algorithms is even easier. These algorithms are not random but they may yield erroneous boundaries, relative to the assumed gold standard. In estimation-theoretic terms, they have zero variance but nonzero bias. If algorithm \( m \) produces boundaries \( \hat{\mathbf{B}}_{jm} \) for patient \( j \), \( \text{Pr}(TC) \) for that algorithm and patient is given by

\[
\hat{\text{Pr}}(TC|\alpha, j, m) = \text{Pr}(TC|D(\hat{\mathbf{B}}_{jm}, \mathbf{G}_{j}), \mathbf{B}_{j}, j).
\]

Again, this expression depends implicitly on \( \alpha \), and a similar expression holds for \( \hat{\text{Pr}}(NTC|\alpha, j, m) \), so a TOC curve specific to algorithm \( m \) and patient \( j \) is readily constructed. Averaging over patients then provides a task-based metric for the particular algorithm.

5. Application to patient motion

Now let’s go back to the general theory from section 2 and see what assumptions should be made for purpose of evaluation of the effect of patient motion rather than segmentation algorithms. In this case we are not concerned with the effects of image noise or random segmentation algorithms, so the PDFs that describe those effects, \( \text{pr}(\mathbf{G}_{j}|j) \) and \( \text{pr}(\hat{\mathbf{B}}_{j}|\mathbf{G}_{j}) \), respectively, can be treated as delta functions in equations (2) and (3). Equivalently, we can simply omit the integrals over \( \mathbf{G}_{j} \) and \( \hat{\mathbf{B}}_{j} \), so that (2) and (3) become

\[
\text{Pr}(TC|\alpha, j) = \int d\mathbf{D}_{j} \text{Pr}(TC|\mathbf{D}_{j}, \mathbf{B}_{j}, j) \text{pr}(\mathbf{D}_{j}|\hat{\mathbf{B}}_{j}, \mathbf{G}_{j}, j),
\]

\[
\text{Pr}(NTC|\alpha, j) = \int d\mathbf{D}_{j} \text{Pr}(NTC|\mathbf{D}_{j}, \mathbf{B}_{j}, j) \text{pr}(\mathbf{D}_{j}|\hat{\mathbf{B}}_{j}, \mathbf{G}_{j}, j).
\]
Another way of expressing these same results is to note that the true boundaries during therapy are some transformation of the original boundaries used in the treatment planning. Thus we write

$$B_j = T_j \hat{B}_j,$$

where $T_j$ is a transformation operator describing the motion and deformation of patient $j$. This transformation is, of course, random, and we describe it conceptually with a PDF $pr(T_j)$. This introduces another integral in (11) and (12), so they become

$$Pr(TC|\alpha, j) = \int dD_j \int dT_j Pr(TC|D_j, B_j, j) pr(D_j|\hat{B}_j, G_j, T_j, j) pr(T_j).$$

(14)

$$Pr(NTC|\alpha, j) = \int dD_j \int dT_j Pr(NTC|D_j, B_j, j) pr(D_j|\hat{B}_j, G_j, T_j, j) pr(T_j).$$

(15)

To get back to a single integral, we assume, as in the segmentation case, that the dose-planning algorithm delivers accurate dose distributions based on the boundaries and CT data input to it. In the segmentation case, this assumption was expressed by (4), repeated here for convenience:

$$pr(D_j|\hat{B}_j, G_j, j) = \delta[D_j - D(\hat{B}_j, G_j)]$$

(4),

where, as above, $D(\hat{B}_j, G_j)$ is the dose distribution (interpreted as a function of continuous spatial variables) returned by the planning algorithm. To apply this assumption to patient motion, however, we must recognize that the patient motion affects not only the boundaries but also the distribution of attenuation and scattering coefficients within the boundaries. With a slight abuse of notation, we will use the same operator $T_j$ to describe the transformation of the CT values, so

$$pr(D_j|\hat{B}_j, G_j, j) = \delta[D_j - D(T_j\hat{B}_j, T_jG_j)].$$

(16)

The interpretation here is that $D(T_j\hat{B}_j, T_jG_j)$ is the dose distribution computed by the planning algorithm with perfect knowledge of the boundaries and attenuation coefficient distribution of patient $j$ after the deformation. In fact, for the purposes of the integrals in (14) and (15), it is rigorously correct to assume this perfect knowledge because we require the PDF $pr(D_j|\hat{B}_j, G_j, T_j, j)$, which is conditional on $T_j$.

Now using (16) in (14) and (15), we obtain

$$Pr(TC|\alpha, j) = \int dT_j Pr(TC|D(T_j\hat{B}_j, T_jG_j)) pr(T_j),$$

(17)

$$Pr(NTC|\alpha, j) = \int dT_j Pr(NTC|D(T_j\hat{B}_j, T_jG_j)) pr(T_j).$$

(18)

As in section 3, the expressions in (17) and (18) suggest a Monte Carlo algorithm that can be implemented with simulated patient motion.

Suppose we begin with a segmented CT image for a particular patient $j$ and we generate $K$ new segmented CT images by randomly sampling different transformation operators $T_j$ according to some model of the motion. The $k^{th}$ such transformation is denoted $T_j^{(k)}$. We plan the therapy on the basis of the untransformed CT and boundary data, but then we run the dose calculation with each of the transformed data sets, generating new dose distributions for each, and input each into the modules that calculate the TC and NTC probabilities.

The Monte Carlo estimate of $Pr(TC|\alpha, j)$ is then

$$\hat{Pr}(TC|\alpha, j) = \frac{1}{K} \sum_{k=1}^{K} Pr(TC|D(T_j^{(k)}\hat{B}_j, T_j^{(k)}G_j)).$$

(19)
where the probability on the right-hand side is interpreted as the output of the TCP module for the $k^{th}$ transformation of the CT and boundary data from patient $j$.

Similarly,

$$\hat{Pr}(N|\alpha, j) = \frac{1}{K} \sum_{k=1}^{K} Pr\left[ N|D(\mathcal{T}_j^{(k)} \hat{B}_j, \mathcal{T}_j^{(k)} G_j) \right].$$

(20)

The Monte Carlo estimates of (19) and (20) are still functions of $\alpha$, so the computation gives one point on the TOC curve. Repeating for several values of $\alpha$ as in an ROC study generates the TOC curve for phantom $j$. Thus one can obtain a TOC curve and AUTOC specific to patient $j$ (with the motion model perhaps reflecting how ‘wiggly’ that patient is). The study can also be repeated for several patients to get an average TOC curve and AUTOC, and these average values can be used to assess the effectiveness of devices and algorithms for measuring and controlling patient motion.

To illustrate the application of the TOC concept to patient motion, assume that a CT image is acquired and used to produce a dose plan. If there were no patient motion, one could produce a TOC curve as described above with overall dose $\alpha$ as the hidden variable (see the ‘no patient motion’ TOC curve in figure 3(a)). To compute the Monte Carlo estimate of $Pr(TC|\alpha, j)$ and $Pr(NTC|\alpha, j)$ we need to generate random samples of the motion operator $\mathcal{T}_j$. The TOC operating points for a fixed $\alpha$ but for many realizations of $\mathcal{T}_j$ is shown as the ‘possible operating points’ in figure 3(a). The smaller this ‘cloud’ of TOC operating points is, the less of an effect patient motion has on the dose plan. When we compute the TOC operating point using (19) and (20) we are averaging over patient motion to create (when $\alpha$ is used as a hidden variable) an average TOC curve (also shown in figure 3(a)). Figure 3(b) shows an alternate description. For every realization of patient motion $\mathcal{T}_j$, there is a separate TOC curve generated by varying $\alpha$. The range of possible TOC curves is illustrated as the shaded region in figure 3(b). The goal of a motion compensation method is to reduce the size of the cloud of possible TOC curves which will, in turn, raise the average TOC curve. Note that the TOC curve with no patient motion does not change so a smaller cloud of possible TOC curves must result in a higher average TOC curve. 

Figure 3. Illustrations of the effect of patient motion on a TOC curve.
So far we have assumed implicitly that the transformation is constant during a treatment, but if the motion model allows a time-varying deformation $T^{(k)}(t)$—and if computer time is no obstacle—we can also define a time-averaged dose distribution as

$$\overline{D}_j^{(k)} = \frac{1}{T} \int_0^T dt D(T^{(k)}(t) \hat{B}_j, T^{(k)}(t) G_j),$$

where $T$ is the treatment time. Of course, $\overline{D}_j^{(k)}$ will be estimated in practice by a finite sum over discrete times rather than a continuous integral, and the dose calculation will have to be run for each time point.

With this modification, (19) becomes

$$\hat{P}_r(TC|\alpha, j) = \frac{1}{K} \sum_{k=1}^K Pr\left[TC|\alpha, \overline{D}_j^{(k)}\right],$$

and similarly for NTCP.

This expression is general enough to include the case where the beam profiles or other treatment parameters vary with time if we assume that the TCP and NTCP depend only on the total dose delivered and not on the dose rate. It could also be modified for the case where the patient was reimaged and the treatment replanned between fractions, though the required computer time increases with each such generalization.

Still another possible generalization is to build the motion model and all of this TOC formalism into the therapy planning process itself. If we have believable TCP and NTCP modules and enough computer power, we can conceivably use AUTOC as the objective function for therapy optimization.

### 6. Application to the use of molecular imaging in therapy planning

So far the only imaging modality we have considered explicitly is CT, but molecular imaging (MI) is also playing an increasingly important role in radiation therapy. For example, PET SPECT and EPR (electron paramagnetic resonance) imaging can be used to delineate the boundaries of hypoxic regions of a tumor (Krohn et al. 2008, Halpern et al. 1994, Rischin et al. 2006, Elas et al. 2008), and MRI sequences sensitive to pH can be used to identify regions of rapid anaerobic growth (Raghunand et al. 2003, Mori et al. 1998, Ward and Balaban 2000). In addition to gross tumor boundaries seen on CT, the boundaries of these subregions require special attention as part of the therapy planning.

The TOC formalism above is readily modified for the case where estimated boundaries obtained from MI are used in the therapy planning, either instead of or in addition to boundaries estimated from CT. For example, if both kinds of boundary estimates are used, and we assume that the noise in the two modalities as well as any errors in the boundary estimates from the two modalities are statistically independent, (5) becomes

$$Pr(TC|\alpha, j) = \int d\hat{B}_j^{CT} \int d\hat{B}_j^{MI} \int dG_j^{CT} \int dG_j^{MI} Pr\left[TC|D(\hat{B}_j^{CT}, \hat{B}_j^{MI}, G_j^{CT}), \hat{B}_j^{CT}, \hat{B}_j^{MI}, G_j^{CT}\right]$$

$$\times pr(\hat{B}_j^{CT}|G_j^{CT}) pr(G_j^{CT}|j) pr(\hat{B}_j^{MI}|G_j^{MI}) pr(G_j^{MI}|j),$$

where we have assumed that $D_j = D(\hat{B}_j^{CT}, \hat{B}_j^{MI}, G_j)$, as in (2) and (5). In other words, the dose distribution is fully determined by the estimated boundaries used as input to the treatment planning and the photon attenuation coefficients derived from CT, but in addition the probability of tumor control depends on the actual boundaries. A result similar to (22) applies to the probability of normal-tissue complications, and that result and (22) are sufficient to compute the TOC curve and AUTOC.
For the purpose of understanding the effect of MI on AUTOC, it is reasonable to assume that the segmentation algorithms used to estimate the boundaries are accurate, so that
\[
\text{Pr}(\hat{B}_j^{CT}|G_j^{CT}) = \delta[\hat{B}_j^{CT} - B_j^{CT}]; \quad \text{Pr}(\hat{B}_j^{MI}|G_j^{MI}) = \delta[\hat{B}_j^{MI} - B_j^{MI}].
\] (23)

With these assumptions, (22) simplifies to
\[
\text{Pr}(TC|\alpha, j) = \int dG_j^{CT} \int dG_j^{MI} \text{Pr}[TC|D_j, B_j^{CT}, B_j^{MI}, G_j, j] \text{Pr}(G_j^{CT}|j) \text{Pr}(G_j^{MI}|j).
\] (24)

Because of the assumption that inaccuracies in boundary estimation are negligible, the only remaining random effects are the noise in the CT and MI images, represented by the two remaining integrals in (22). Of these two effects, only the CT noise affects \(\text{Pr}(TC|\alpha, j)\), and this effect arises only because the CT noise can affect the estimated photon attenuation coefficients that go into the planning algorithm. If we make the additional assumption that error in the attenuation coefficients can be neglected for the purpose of studying the impact of MI on AUTOC, we can omit these integrals and write
\[
\text{Pr}(TC|\alpha, j) = \text{Pr}[TC|\alpha D_{0j}, B_j^{CT}, B_j^{MI}, G_j, j],
\] (25)

where we have used (1) to replace \(D_j\) with \(\alpha D_{0j}\). Similarly, we can write
\[
\text{Pr}(NTC|\alpha, j) = \text{Pr}[NTC|\alpha D_{0j}, B_j^{CT}, B_j^{MI}, G_j, j].
\] (26)

With these results, it is now trivial to construct a TOC curve for patient \(j\) and compute AUTOC\(_j\). We merely pick some nominal parameter settings and use the segmented CT and MI images in the treatment-planning program to get a baseline dose distribution \(D_{0j}\). We can scale this dose by \(\alpha\) to get \(\alpha D_{0j}\), which is then fed into the TCP and NTCP modules to yield \(\text{Pr}(TC|\alpha, j)\) and \(\text{Pr}(NTC|\alpha, j)\), respectively, for the particular patient \(j\), the patient-specific treatment plan and the given value of \(\alpha\); varying \(\alpha\) in this procedure then traces out the TOC curve with as many points as desired, and AUTOC for this patient and treatment plan is easily obtained by numerical integration.

If one wants to compare this AUTOC to the one that would be obtained without the use of the MI data, all that is required is to ignore the MI data and assume that the MI boundaries are the same as those obtained with the CT data.

7. Application to internal radiotherapy with the sodium iodide symporter

So far we have discussed only external-beam radiation therapy, but there is also considerable interest in methods where the radiation source is inside the body, as in brachytherapy, radioimmunotherapy and radiation therapy mediated by the human sodium iodide symporter gene, hNIS, which is the main topic of this section.

The hNIS protein (Trujillo et al 2010, Gaut et al 2004, Dwyer et al 2005, Spitzweg et al 2001) is used by the human thyroid gland to accumulate iodine from the blood stream as a substrate for synthesizing hormones such tetraiodothyronine (thyroxine). This highly efficient uptake of iodine is exploited through the use of radiolabeled iodine for both imaging and therapy of thyroid disease (Dohan et al 2003). Although initially believed to be thyroid-specific, NIS detection in other tissues raised the possibility for using this effective and safe modality for the treatment of other cancer types. However, native levels of NIS in extrathyroidal tissues are not sufficient to mediate effective treatment (Ryan et al 2011). As a result of this, many studies have investigated the potential for introduction of NIS expression to support imaging and treatment of a variety of cancer types. Broadly, these studies are based on (1) attempts to upregulate native NIS expression through demethylation or stimulation with regulating agents (Provenzano et al 2007, Unterholzner et al 2006), (2) direct introduction of NIS transgene
expression into cancer cells using viral (detailed below) or non-viral vectors (Klutz et al 2009), or (3) the ‘Trojan Horse’ approach, where cellular vehicles, such as mesenchymal stem cells (MSCs), are engineered to express NIS, and then home to the tumor site for delivery of therapy (Dwyer et al 2011, Knoop et al 2011).

Successful migration of the MSCs to the tumor can be verified by SPECT imaging with $^{123}$I or $^{99m}$Tc-pertechnetate, just as if the tumor were a human thyroid, and then the beta emitter $^{131}$I can be used for therapy, just as in clinical thyroid ablation therapy. Because $^{131}$I and $^{123}$I are chemically identical, the therapy isotope and the imaging isotope must have the same spatial distribution.

Another way of using hNIS for internal radiation therapy is to inject genetically modified viruses such as adeno, measles, or other viral hosts that contain and express the hNIS gene. These viral constructs and engineered NIS expression can be made tumor- or tissue-specific by way of transcriptional targeting with tumor-specific promoters, thereby sparing non-target tissues from viral or radioiodide toxicities (Spitzweg et al 2001, Trujillo et al 2010). Replication of the virus results in cytolysis, and this can also be controlled genetically through the use of tumor-specific promoter activity driving viral genes that are critical for assembly and release of active viral particles. The term ‘radiovirotherapy’ describes the synergistic results seen when cytolysis by use of a replication competent virus is also accompanied by NIS expression and radioiodine therapy (Walrand et al 2012, Dingli et al 2003, Trujillo et al 2010, Dwyer et al 2005).

For hNIS-based radiotherapy with $^{131}$I, the image dataset must include both CT, to obtain the anatomical information needed for dose calculation, and a SPECT image with either pertechnetate or $^{123}$I. Thus $G_j$ refers to the set, $\{G_j^{\text{CT}}, G_j^{\text{SPECT}}\}$.

Given the imaging data for patient $j$, one must then estimate the distribution of the radiation dose $D_j$ and from that compute PTC and PNTC. The most general integral expression for the probability of tumor control is [cf (2)]

$$\Pr(TC|\alpha, j) = \int dD_j \int dA_j^{Tx} \int dG_j \Pr(TC|D_j, j) \Pr(D_j|A_j^{Tx}) \Pr(A_j^{Tx}|G_j) \Pr(G_j|j),$$  \hspace{1cm} (27)

where $A_j^{Tx}$ is the activity per unit volume of the therapy isotope, e.g., in MBq cm$^{-3}$. The dependence on $\alpha$ is discussed below.

In (27), $\Pr(G_j|j)$ describes the noise in the image data; $\Pr(A_j^{Tx}|G_j)$ describes any errors or randomness in estimating the activity per unit volume of the therapy isotope from the SPECT data, and $\Pr(D_j|A_j^{Tx})$ similarly describes the inaccuracies or randomness in relating the activity per unit volume of the therapy isotope to the dose distribution (in grays). This step will be random as well as inaccurate if it is performed by a Monte Carlo algorithm (El Naqa et al 2012). The factor $\Pr(TC|D_j, j)$ in (27) is just the output of the same tumor-response code as is used for external-beam therapy.

Once again, we can make some reasonable simplifying assumptions. First, if we neglect inaccuracy and randomness in the modules for computing TCP and NTCP, as we did in (4), we can omit the integral over $D_j$ just by regarding $D_j$ as the module output, $D(A_j^{Tx})$.

The next factor to consider is the noise in the CT and SPECT images. If the noise is negligible, we can say formally that $\Pr(G_j|j)$ is a delta function, which is the same as simply omitting the integral over $G_j$ as in section 5.

The factor $\Pr(A_j^{Tx}|G_j)$ can be written as $\Pr(A_j^{Tx}|G_j^{\text{SPECT}})$ if the CT image is not used in the SPECT reconstruction, and for hNIS therapy with $^{125}$I (but not for hNIS with pertechnetate or for radioimmunotherapy with surrogate SPECT agents), we have that

$$A_j^{Tx} = \alpha A_j^{\text{SPECT}},$$  \hspace{1cm} (28)

where $\alpha$ is the ratio of injected $^{131}$I activity to injected $^{125}$I activity.
Thus pr(A_j^T | G_j^{SPECT}) describes any inaccuracy in measuring the ratio of injected activities as well as the error with estimating activity from a SPECT image. The estimation error arises from three main sources: collimator blur and scattering and attenuation of radiation in the patient’s body. Retaining this factor and the associated integral for now, we have

$$\text{Pr}(TC(\alpha, j)) = \int dA_j^T \text{Pr}[TC|D(A_j^T)] \text{pr}(A_j^T | G_j^{SPECT}),$$  

(29)

and, of course, similarly for Pr(NTC(\alpha, j)).

If the objective of an image-quality study is to assess the impact of SPECT hardware or reconstruction algorithms on therapeutic efficacy as expressed by AUTOC, the integral in (29) can be evaluated by detailed, realistic Monte Carlo simulations. The index j in this case must refer to a particular simulated patient anatomy and pattern of radioiodine uptake. Averaging over simulated patients then gives an average TOC curve and an average AUTOC that can be used as figure of merit for image quality based on the efficacy of the hNIS-mediated therapy.

If, on the other hand, the quantitative accuracy of the SPECT reconstruction algorithm and the measurement of the activity ratio have been established, and the goal is to compute the TOC curve for a particular patient, we can write

$$\text{Pr}(TC(\alpha, j)) = \text{Pr}(TC|D(\alpha A_j^{SPECT})).$$  

(30)

Thus, after calibration and validation studies, it should be possible to use the TCP/NTCP modules and the resulting TOC and AUTOC to assess and optimize the therapy for each patient.

8. Discussion and conclusions

All of the expressions given above for Pr(TC) and Pr(NTC) and their estimates are specific to a particular patient (or phantom), so we can generate a patient-specific TOC curve. For detection studies, by contrast, an ROC curve is defined only for a population of patients. To appreciate why this difference arises, we can go back to figure 1 which plots the probability of a beneficial outcome (tumor control) and an adverse outcome (normal-tissue complication) against a controllable variable (overall dose). The curves in figure 1 would apply directly to lesion detection if we simply interpreted the beneficial outcome as a true-positive detection, the adverse outcome as a false-positive detection and the controllable variable as the observer’s decision threshold.

For a given patient, we can define the TOC curve, the segmentation-averaged TOC curve, the motion-averaged TOC curve, and many more. In addition, we can then average these curves over patients to generate overall TOC curves that include patient variability. Thus, the TOC curve is a very general framework that can be used to assess the effect of these controls on therapeutic efficacy. One of the main differences between TOC and conventional ROC curves, is in the interpretation of the probabilities of adverse and beneficial outcomes. In the detection problem these probabilities are defined for some hypothetical ensemble of patients presenting for the same diagnostic study, but in the therapeutic problem the TCP and NTCP modules provide a way of computing the probabilities from knowledge of the actual delivered dose distribution and the true boundaries of the tumor and sensitive normal tissues.

It is also interesting to study what TOC curves teach us about ROC curves. Using a similar framework as we describe in this paper, it is possible to generate patient-specific ROC curves that assess the effects of, for example, image noise on the ability of the observer to detect an abnormality, in the well-known signal-known-exactly, background-known-exactly (SKE/BKE) analysis. In other words, knowing the background and signal exactly is akin to
imaging the same patient every time—only the image noise varies. To put it another way, the ROC curve for an SKE/BKE problem is very much like a TOC curve for a particular patient.

For purposes of image-quality assessment, the AUTOC figure of merit must be averaged over some group of patients or other subjects, including possibly simulated patients or animal subjects. Image quality is necessarily a statistical measure and cannot be determined based on the performance on a single patient. This immediately raises the question of assessing the systematic and random errors that affect the determination of AUTOC. Among systematic errors, we must consider the TCP and NTCP modules themselves, which can be inaccurate because of oversimplified radiobiological models or uncertainty in the parameters that go into them. For the purpose of image-quality assessment, however, it is straightforward to see to what extent the modules influence our conclusions. Suppose, for example, that we are using AUTOC to compare segmentation algorithms, as suggested in Barrett et al (2010). Once we have the segmented images and have run the dose-planning algorithm for each, we can readily use many different TCP and NTCP modules and multiple choices of radiobiological parameters in each. If it turns out that the AUTOCs for segmentation algorithm A are consistently better than those for algorithm B, we can be confident in asserting that A is the method of choice for radiation therapy planning. On the other hand, we can also use the same segmented images and dose plans to generate TOC curves for different forms of normal-tissue complication; it may turn out that the choice of segmentation algorithm is different for different complications.

The most important random effect that arises in both the TOC and ROC approaches to task-based assessment of image quality is patient-to-patient variability. How many patients must one use to get a specified accuracy in estimates of the AUCs, and how can one assess the statistical significance of differences between the curves? This is a huge area of research in ROC analysis, stemming from seminal work by Charles Metz, Robert Wagner, Don Dorfman and many others. Similar analyses for TOC would be very useful.

Finally, though we have characterized the TOC curve as akin to an ROC for an individual patient, we can step back from the individual patient analysis and use a population-based TOC analysis to determine the marginal benefit provided by additional or improved imaging information. For example, we can compare AUTOC with and without MI information. This comparison can be carried out in the context of a cost-benefit analysis where the actual dollar cost of the molecular image is balanced against the reduced medical cost of improved cancer control and the reduction of the cost of treatment of normal tissue complication. In this context the TOC approach provides an algorithmic and relatively bias-free means by which to assess the ultimate social usefulness of an imaging modality. Studies might begin with animal model systems to which local image-guided therapies are applied.

Acknowledgments

This research was supported by the National Institutes of Health under grants R37 EB000803 and P41 EB002035 at the University of Arizona; P41 EB002034 and R01 CA0988575 at the University of Chicago; and P50 CA91956 at the Mayo Clinic, which was also supported by Minnesota Partnership for Biotechnology and Medical Genomics, Grant 08-06.

We wish to thank Luca Caucci and Craig Abbey for careful reading of the manuscript and thoughtful comments.

References

Moore D H II and Mendelsohn M L 1972 Optimal treatment levels in cancer therapy Cancer 30 97–106
Romeijn H E and Dempsey J F 2008 Intensity modulated radiation therapy treatment plan optimization TOP 16 215–43
Trujillo M A, Oneal M, McDonough S, Qin R and Morris J C 2010 A probasin promoter, conditionally replicating adenovirus that expresses the sodium iodide symporter (NIS) for radiovirotherapy of prostate cancer Gene Ther. 17 1325–32