Methodological issues in radiation dose–volume outcome analyses: Summary of a joint AAPM/NIH workshop

Joseph O. Deasy[a)]
Washington University School of Medicine, St. Louis, Missouri 63110

Andrzej Niemierko
Massachusetts General Hospital & Harvard Medical School, Boston, Massachusetts 02114

Donald Herbert
College of Medicine, University of South Alabama, Mobile Alabama 36617

Di Yan
William Beaumont Hospital, Royal Oak, Michigan 48073

Andrew Jackson
Memorial Sloan Kettering Cancer Center, New York, New York 10021

Randall K. Ten Haken
University of Michigan, Ann Arbor, Michigan 48109

Mark Langer
Indiana University, Indianapolis, Indiana 46202

Steve Sapareto
Good Samaritan Regional Medical Center, Phoenix, Arizona 85006

(Received 8 January 2002; accepted for publication 28 June 2002; published 26 August 2002)

This report represents a summary of presentations at a joint workshop of the National Institutes of Health and the American Association of Physicists in Medicine (AAPM). Current methodological issues in dose–volume modeling are addressed here from several different perspectives. Areas of emphasis include (a) basic modeling issues including the equivalent uniform dose framework and the bootstrap method, (b) issues in the valid use of statistics, including the need for meta-analysis, (c) issues in dealing with organ deformation and its effects on treatment response, (d) evidence for volume effects for rectal complications, (e) the use of volume effect data in liver and lung as a basis for dose escalation studies, and (f) implications of uncertainties in volume effect knowledge on optimized treatment planning. Taken together, these approaches to studying volume effects describe many implications for the development and use of this information in radiation oncology practice. Areas of significant interest for further research include the meta-analysis of clinical data; inter-institutional pooled data analyses of volume effects; analyses of the uncertainties in outcome prediction models, minimal parameter number outcome models for ranking treatment plans (e.g., equivalent uniform dose); incorporation of the effect of motion in the outcome prediction; dose-escalation/isorisk protocols based on outcome models; the use of functional imaging to study radioresponse; and the need for further small animal tumor control probability/normal tissue complication probability studies.
[DOI: 10.1118/1.1501473]

Key words: outcomes analysis, TCP, NTCP, volume effects

I. INTRODUCTION

The rapid advance and acceptance of three-dimensional conformal radiation therapy and intensity modulated radiation therapy is revolutionizing the delivery of radiation therapy. Despite this development, the ability to evaluate and compare the resulting highly complex dose distributions with respect to expected clinical outcomes has lagged far behind. One major factor has been the lack of useful data available to assess and develop useful predictive models. A second major factor is the need for new plan-ranking methods that correlate inherently multidimensional dose distribution data with outcomes. This report represents a summary of presentations at a joint workshop of the National Institutes of Health and the American Association of Physicists in Medicine (AAPM) held in Bethesda, Maryland, 22 October 1999. It is the first publication of the Biological Effects Committee Task Group 8, Volumetric Tissue Response Models in Radiation Oncology, organized by the Biological Effects Committee of the AAPM. Volume effect issues are addressed here from several different perspectives. The topics discussed include (a) basic modeling issues including the equivalent uniform dose framework and the bootstrap method, (b) issues in the valid use of statistics, including the need for meta-analysis, (c) issues in dealing with organ deformation and its effects on

treatment response, (d) evidence for volume effects in a clinical site (rectum), (e) the use of volume effect data in liver and lung as a basis for dose escalation studies, and (f) implications of uncertainties in volume effect knowledge on clinical practice. The presentations and references were updated for publication. Taken together, these approaches to studying volume effects describe many implications for the development and use of this information in radiation oncology practice.

II. MODELING THE EFFECT OF INHOMOGENEOUS DOSE DISTRIBUTIONS

Andrzej Niemierko

Dose distributions in radiation therapy are inherently inhomogeneous. This is due to the heterogeneity and geometry of the irradiated tissues, and the physics of radiation beams designed to deliver high dose to the tumor volume while maximally sparing surrounding normal structures. The advent of three-dimensional treatment planning and, more recently, Intensity Modulated Radiation Therapy (IMRT) created tools for planning, and means for delivering, optimized dose distributions tailored to the patient's geometry. The IMRT dose distributions are often significantly different from the corresponding dose distributions obtained using nonmodulated fields. Even when treatment planning is performed according to a specific protocol the IMRT dose distributions can be also substantially different from case to case, as a result of much greater flexibility available in IMRT for plan individualization. However, this flexibility has also some drawbacks. For example, optimized IMRT target dose distributions sometimes have relatively deep cold spots, usually on the periphery of the target volume. The clinical experience with such dose distributions is limited, which makes a quantitative comparison of rival plans more difficult.

To take full advantage of the available technology and to optimize and individualize radiation treatment, one needs tools to assess the potential clinical outcome of any radiation treatment. That is, one needs tools that quantify the probability of an end point of interest [e.g., Tumor Control Probability (TCP), Normal Tissue Complication Probability (NTCP), or five-year recurrence-free survival] as a function of delivered radiation treatment (i.e., the volumetric and temporal distribution of dose) and as a function of the most meaningful characteristics of the irradiated tissues and organs (i.e., their geometry and biology). Of course, the response of any biological systems to radiation is complex and multifaceted. It is unfeasible to account for all the processes contributing to the observed outcomes. However, we can try to identify and model the dominant ones.

Since radiation directly kills cells, much effort has been expended to investigate cell survival characteristics and to develop models of tissue response to radiation based on the so-called "target cell hypothesis." [1,2] By definition, the target cells are those cells that are capable of replication and regeneration of the tissue. Hence, it is hypothesized that tissue and organ response to radiation can be described in terms of the survival probability of the target cells. Tissues and organs differ in their target cell survival characteristics and their organization. Consequently, tissues and organs differ in their tolerances and their response to altered dose distributions and fractionation. These differences have important implications for treatment planning and for optimization of radiation therapy. For example, IMRT has the capability of "shaping" dose distribution by using a large number of individually weighted (modulated) "beamlets." This capability allows partial sparing of the sensitive critical structures surrounding the target volume while still delivering a high dose to the tumor. Obviously, to optimally distribute unavoidable dose to critical normal structures and to quantify the difference between competing dose distributions one needs quantitative models of dose–volume and dose–response relationships for all structures of interest.

The fundamental requirement for any dose optimization process is to be able to rank the competing dose distributions. That is, to be able to tell which dose distribution is better. Note that for plan ranking it is unnecessary to know by how much one plan is better than the other. The physical dose is an easy-to-understand and convenient metameter of radiation treatment planning. As a rule, more dose to the target volume and less dose to the critical structures is considered to be better. However, because dose distributions are three-dimensional inhomogeneous objects they cannot be ranked directly. That is, it is meaningless to ask what dose was actually delivered to a structure of interest if the dose distribution within that structure is nonuniform.

A concept of Equivalent Uniform Dose (EUD) was developed for more precise reporting, analyzing, and ranking dose distributions actually delivered to irradiated organs and volumes of interest.[3] The concept of EUD assumes that any two dose distributions are equivalent if they cause the same radiobiological effect of interest. For example, for tumors it is often postulated that the probability of local control is determined by the expected number of surviving clonogens, according to Poisson statistics. Therefore, it can be shown that for any inhomogeneous target dose distribution the EUD is calculated as follows:

$$\mathrm{EUD} = 2\ \mathrm{Gy}\,\frac{\ln\left[\frac{1}{N}\Sigma_{i=1}^{N}(\mathrm{SF}_2)^{D_i/2\ \mathrm{Gy}}\right]}{\ln(\mathrm{SF}_2)}, \tag{1}$$

where SF_2 is the fraction of clonogens surviving a dose of 2 Gy and the sum is taken over all N dose calculation points within the target volume and D_i is the dose at point i. Using the formalism of the Linear–Quadratic model but suppressing the quadratic component for simplicity here (see Ref. 3 for the more complex equation which results if the quadratic component is included) the EUD is calculated as follows:

$$\mathrm{EUD} = -\frac{\ln\left[\frac{1}{N}\Sigma_{i=1}^{N}\exp(-\alpha D_i)\right]}{\alpha}, \tag{2}$$

where α is the initial slope of the survival curve.

It is important to recognize that the EUD formalism expressed in Eqs. (1) and (2) represents a situation where both dose distributions, i.e., the original inhomogeneous one and

the equivalent uniform one, are delivered with the same number of fractions with identical interfractional interval and identical overall time. More generally, all nonvolumetric parameters of the treatment plan are constant.

There are several important consequences of the clonogen survival-based EUD concept. (1) The EUD for tumors is always bounded by the minimum target dose and the mean target dose. (2) For greater dose inhomogeneity the EUD tends toward the minimum target dose; for lesser inhomogeneity it tends toward the mean target dose. (3) The detrimental effect of a cold spot is more pronounced for tumors characterized by more sensitive clonogens (i.e., lower SF_2 or higher α). Reference 3 contains logical extensions of this EUD model to account for absolute volume effect, dose per fraction and proliferation effects, and possible nonuniform spatial distribution of clonogens. It should be noticed that for treatment plan optimization, the simple EUD model presented here with the parameter SF_2 set to 0.5 is often sufficient. The EUD concept has been demonstrated to be useful for clinical data analysis[4] and for optimization of IMRT.[5]

The EUD model can be extended to normal tissues based on the concept of Functional Sub-Units (FSUs).[6,7] It is hypothesized that for most normal structures the outcome is determined by the number or the fraction of the FSUs that survive the treatment. Most normal organs exhibit some sort of functional reserve that allows the organ to maintain its functionality, even when some fraction of that organ is completely destroyed. Consequently, the FSU-based EUD model exhibits the strongest dependence on the subvolume of the irradiated organ that receives no or very little dose. The EUD for normal structures is usually bounded by the mean and the maximum dose to the structure of interest. This relationship depends upon the survival characteristics of the hypothetical FSUs and their organization. For example, for structures with a "serial" organization (spinal cord and esophagus are often given as examples), the EUD is closer to the maximum dose. For structures with a "parallel" organization (e.g., lung, liver, kidney) the EUD is closer to the mean dose.[8]

The effects of dose inhomogeneity can be also accounted for by using phenomenological models, such as the dose–volume histogram reduction schemes based on the power law.[8–10] One can argue that it is practically impossible to develop a comprehensive biological model of tissue response to radiation and therefore these phenomenological approaches may present a better (or at least less model-dependent) choice. Based on the power law the EUD can be calculated as follows:

$$\mathrm{EUD}=\left(\frac{1}{N}\sum_{i=1}^{N} D_i^a\right)^{1/a}. \tag{3}$$

This EUD formula has been proposed for both tumors and normal structures.[8] The model parameter "a" is negative for tumors and positive for normal structures. It is easy to see that for large negative "a" the EUD tends to the minimum dose and for large positive "a" the EUD tends to the maximum dose. For "a" equal to one the EUD represents the average dose. The parameter "a" is tissue and end-point specific and the value of "a" needs to be obtained by fitting an EUD-based dose-response model to clinical data.

It is difficult to differentiate and validate either biology-based or phenomenology-based models using statistically adequate methods because the available clinical data are usually incomplete and too weak. On the other hand, because the data are weak it is often possible and easy to "fit" them with a wrong model. That is, the data may be too weak to reject a bad model. Nevertheless, if cautiously used either model can provide helpful quantitative measures for evaluating dose distributions.

A logical consequence of the concept of EUD is to use it as a sole argument of the models of tissue response to radiation. This is because all the relevant effects (volumetric, temporal, and others) can be accommodated within the concept of EUD. Hence, the probability of effect to an end point of interest can be estimated using the corresponding EUD and any convenient dose-response function such as Logit, Probit, Poisson, or exponential.[11–13]

III. USEFUL MODELS AND ADEQUATE METAPHORS

Donald Herbert

"*All models are wrong, but some are useful.*" G.E.P. Box, 1979.[14]

Quality of Care and the "*Ethics of Belief*:" In the field of radiation oncology, AAPM Report 43 (1993[15]) describes in detail, "*...the statistical inadequacy of many historical and contemporary studies which have been accepted as the basis for the construction of models used in clinical practice. Use of these* ***flawed models*** *may well have retarded progress in the past and, more importantly, still threatens to inhibit future advances in the practice of radiation oncology.*"—L. Cohen, 1993,[16] a finding with implications for the work of Task Group 8 (see below).

These findings are quite consistent with other evaluations of the quality of the medical literature, e.g., "*Serious widespread problems exist in the clinical literature...the average practitioner will find relatively few journal articles that are scientifically sound in terms of reporting usable data and providing even moderately strong support for their inferences.*" [17] Such findings are disturbing in view of the recent IOM report on the quality of health care.[18] The National Cancer Policy Board, in its 1999 publication, "*Ensuring Quality Cancer Care,*" [19] used the IOM 1998 definition of quality cancer care: "*Health care can be judged as good to the extent that it increases the likelihood of* ***desired health outcomes*** *and is* ***consistent with current professional knowledge***." This has two immediate implications for the work of AAPM Bioeffects Committee Task Group 8: First, assessments of "*desired health outcomes,*" should include psychometrically valid measures of the functional status of the patient such as is provided by the MOS SF-36 questionnaire.[20] The 1989 Patient Outcome Research Act[21] mandated that measures of "*functional status and well-being and patient satisfaction*" be included with traditional mea-

sures of survival, clinical end points and disease. Functional status measurements are important not only as "*outcomes*" for payers but also as covariates in models predicting clinical "*outcomes*." Second, some investigators have incorporated the LQ model into IMRT models. However, the LQ model is inconsistent with "*current professional knowledge*:" AAPM Report 43 holds the validity of the LQ model to be very problematic. (All models are wrong, but some are awful.) Moreover, even if the validity of the model were less of a problem, the estimates of the parameters, α, β, and their ratio, α/β, are encumbered with large uncertainties, both random and systematic [e.g., the ratio of estimates, est(α)/est(β), is a biased estimate of the ratio, α/β], since, as remarked in AAPM Report 43, they tend to be constructed by "*dubious methods from questionable data*." [15] This is, of course, "*A major problem in many areas...experts often hold strong beliefs that have no evidential foundation. ... A key problem about the ethics of belief is that people may believe strongly in something they know little about and have little evidence for,*" F. Mosteller, 1999.[22]

Meta-analysis: "*Meta-analysis is going to revolutionize how the sciences, especially medicine, handle data. And it's going to be the way many arguments will be ended*." T. Chalmers, 1991.[23] Meta-analysis is a statistical method for combining the information residing in the summary statistics, say odds ratios and their standard errors, from a group of independent studies, e.g., clinical trials, of a given treatment or diagnostic procedure. In biomedical and epidemiological investigations it is usually undertaken to achieve one of two aims: (1) obtaining pooled point and interval estimates of an overall "*true treatment effect*." (2) "*borrowing strength*" across a group of studies to obtain improved point and interval estimates of the effect in each study, i.e., improved study-specific estimates.[24] There are four models of meta-analysis: fixed effects, random effects, mixed effects, and Stein effects. The first three models are used to achieve the first aim. The last two are used to achieve the second aim. The Stein, or empirical Bayes model, like most Bayes models, provides a useful answer to the extrapolation problem often described as *Bernard's dilemma*: The response of the "*average*" patient to therapy is not necessarily the response of the patient being treated.

However, although meta-analysis was initially introduced as a bridge across the gap between undersized controlled trials and the treatment of patients, given the current state of the literature remarked above, it may be equally (or more) useful as a filter: "*Meta-analysis ... spotlights the ... grave defects in the majority of original research that only come to light if an attempt is made to combine the data with that of other research.*" T. Chalmers, 1987.[25] (Such findings determined the direction of AAPM Report 43.)

Validated reviews: As a remedy for the poor quality of the medical literature described above, it is recommended that the "*average practitioner seek reviews based on validated research sources*." [17] The best such reviews are those of the Cochrane Collaboration. The Cochrane Collaboration is a nonprofit international organization that produces high quality systematic reviews (where possible meta-analyses) of all studies, both published and unpublished, of every sort of health care intervention.[26] These are subjected to rigorous evaluation and regularly updated. The results of these reviews are regularly disseminated electronically both on CD ROM and via the Internet. Physicians and scientists seeking the best evidence either to guide their practice or to design their studies will more likely find it in the systematic reviews of the Cochrane Collaboration than in the unaggregated and "*unfiltered*" studies listed on MEDLINE or EMBASE. Naylor has remarked that, "*... the Cochrane Collaboration is an enterprise that rivals the Human Genome Project in the potential implications for modern medicine*." [27]

Clinical databases: Clinical databases are a crucial component in the useful extrapolation of the results of clinical trials to local clinical practice since (1) "*Clinical trials are typically done under carefully controlled circumstances in a university setting and do not reflect the use of a technology in community practice*" [28] *and* (2) "*Patients in a trial are* **not** *to be regarded as a random sample from some population.*" [29] Indeed, "*Data on the quality of community practice by hospital and physician are essential for making proper use of data from randomized trials ... the quality with which each arm of the trial may be carried out in the community may overwhelm any differences between alternative therapies that are demonstrated in the trial itself.*" R. Brook, 1993.[30] (Unfortunately, at present, most clinical databases are inadequate to support useful extrapolations.)

Databases also should be considered for use both as a supplement to randomized clinical trials in *ensemble-of-studies* models, as in the meta-analytic procedure of Cross-Design Synthesis[31] and in Response Surface Designs (see below) and as an alternative to them. In the usual meta-analysis, it is common to pool the summary statistics (e.g., the odds ratio and its standard error) of a set of studies encumbered by common weaknesses. In the case of clinical trials these may be high levels of type I and type II errors. In Cross-Design Synthesis the summary statistics of a set of studies with complementary weaknesses are pooled. Thus, clinical trial results with characteristic low external validity (since patients in the trial are not randomly selected)[29] are pooled with database results with characteristic low internal validity (since rival treatments are not randomly assigned).

Volume effects: Typically there are four categorical **outcomes** to any treatment of a cancer patient: These are the compound events, E_1 and $\overline{E_2}$, $\overline{E_1}$ and $\overline{E_2}$, E_1 and E_2, $\overline{E_1}$ and E_2, where E_1 is ablation of tumor, E_2 is the complication of normal tissue ("*No side effect, no effect*") and the overbar identifies the complementary event. Then treatment success is the event $S=E_1$ and $\overline{E_2}$. Treatment failure is the complementary event $\overline{S}=E_1$ and E_2 or $\overline{E_1}$ and E_2 or $\overline{E_1}$ and $\overline{E_2}$.[32–34] It is the joint probabilities, say $P(E_1 \text{ and } \overline{E_2})$, of compound events that are of interest and since the elementary events (E_1, E_2, etc.) are not independent, a bivariate probit dose-response model[35] is most useful. Models of the joint probability of the occurrence of compound events are

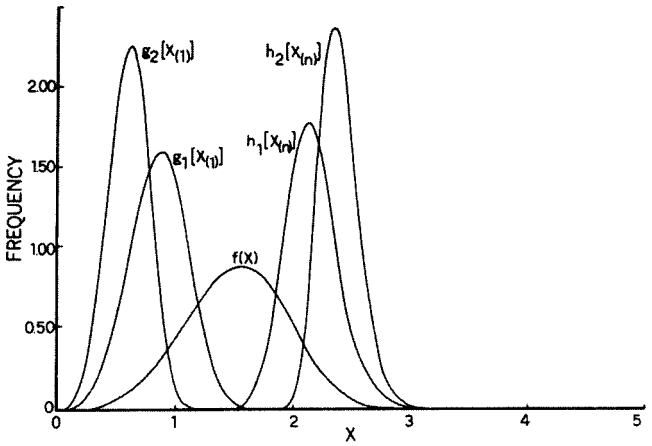

FIG. 1. A superposition of the density funcion of a two-parameter (location and scale) Weibull distributions, $f(x)$, and the density functions of the smallest, $g_i(x_{(1)})$, and largest $h_i(x_{(n)})$, values of x in samples of size n from the parent distribution, $f(x)$. The mode of the distribution of smallest (largest) values decreases (increases) with sample size n, e.g., n is larger for $g_2(x_{(1)})$ than for $g_1(x_{(1)})$.

useful in several endeavors. Two of these are (1) the elucidation of "*volume effects*" and (2) the design of clinical trials.

Describing the outcome of treatment in terms of the compound events $S=(E_1$ and $\overline{E_2})$ provides a useful insight into two of the effects of volume of irradiated tissue on the probability of occurrence of this event. Herbert[34] (1984) showed that his 1983 "*extreme value*" model[33] of volume effects in radiation oncology (based on the Statistical Theory of Extreme Values that has proved so useful in explaining the effects of a cognate "*size effect*" on the strength of materials in engineering studies) predicts (1) that the probability of uncomplicated control of a tumor decreases as the volume increases and (2) that the slopes of the respective dose response curves for ablation of tumor and complication of normal tissue increase as the volume of irradiated tissues increase. Empirical observation strongly suggests that radiation complication in normal tissue is a smallest value phenomenon and that the recurrence of a tumor is a largest value phenomenon, i.e., "*chain*" (series connection) and "*rope*" (parallel connection) models, respectively.[33]

Figure 1 shows the density functions of smallest and largest values of a two parameter Weibull distribution.[34] It shows that as the sample size (volume) increases, the separation between the respective density functions increases: the smallest values get smaller, the largest values get larger, exactly what would be expected if x were a measure of the "*breaking strength*" of the tissue, either normal or tumor. This is the level of ambient "*stress*" at which the functional and/or structural integrity of the tissue fails. For example, for tumor tissue the "*stress*" may be the immunocompetence of the host, cytotoxic drugs, etc.; for normal tissue, the "*stress*" may be physical trauma such as a blow, low temperatures, etc. Figure 2 shows a superposition of the dose-response curves for ablation of tumor, $P(E_1|x_1)$, and necrosis of normal tissue, $P(E_2|x_1)$, at $T=45$ days and $N=33$ fractions for target volumes V_1 (solid) and $2V_1=V_2$ (dashed) in carcinoma of the oropharynx.[21] Note that the *location* and *scale* measures of both dose-response curves are functions of the target volume, V. Note that the *location* of $P(E_1|x_1)$ is shifted to *lower* doses as V increases. Note that the *slope* of each dose-response *increases* as V increases.

Response surface designs: Clinical trials are commonly designed to test hypotheses on the levels of a response variable, e.g., survival, in comparing two or more interventions. However, in Response Surface Designs (RSD), one is interested, instead, in finding, by a Steepest Ascent method, the set of values of the treatment variables that yield a maximum response, typically, the maximum probability of uncomplicated control of disease, a treatment success: $P(S)$. RSD is an iterative, *evolutionary*, extrapolation procedure that constructs polynomial models of second order in the treatment variables ("*useful models*") to a set of dose-response data. It identifies the *local* nature of the resulting surface (ridge, maximum, saddle, etc.) by the sign and size of the eigenvalues of the canonical quadratic form of the surface and if the eigenvalues are not all negative, and of roughly the same size, the corresponding eigenvectors give an estimate of the direction to the maximum.[35] (See Herbert 1997[36] for a heuristic exposition of clinical RSDs.) Figure 3 gives the layout and analysis of a typical RSD in two variables: elapsed time T and total dose D of radiation for treatment of cancer of the oropharynx. The dashed lines identify a family of isoeffect curves for $P(S)$, the probability of uncomplicated control. The constructed response surface is an ellipse; the signs and relative sizes of the eigenvalues suggest that the response surface in this region is a "*ridge*" rising in a direction given by the eigenvector shown.

Data mining: Randomized controlled clinical trials are

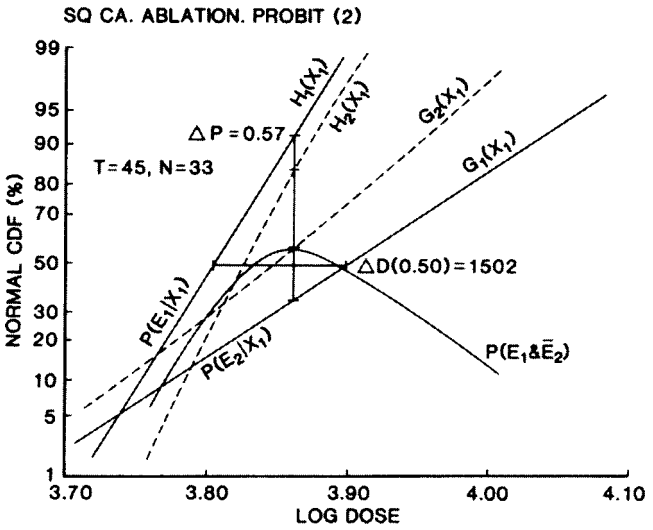

FIG. 2. A superposition of the dose–response curves for ablation of the tumor, $P(\overline{E_1}|x_1)$, and necrosis of normal tissue, $P(E_2|x_1)$, at $T=45$ days and $N=33$ fractions for traget volumes V_1 (solid) and $2\ V_1=V_2$ (dashed). Note that the location of the curve $P(E_2|x_1)$ is shifted to higher levels of dose $x_1(=\log D)$ and $P(E_2|x_1)$ is shifted to lower levels of dose x_1. The slope of both dose–responese curves is also increased as the volume, V, increases.

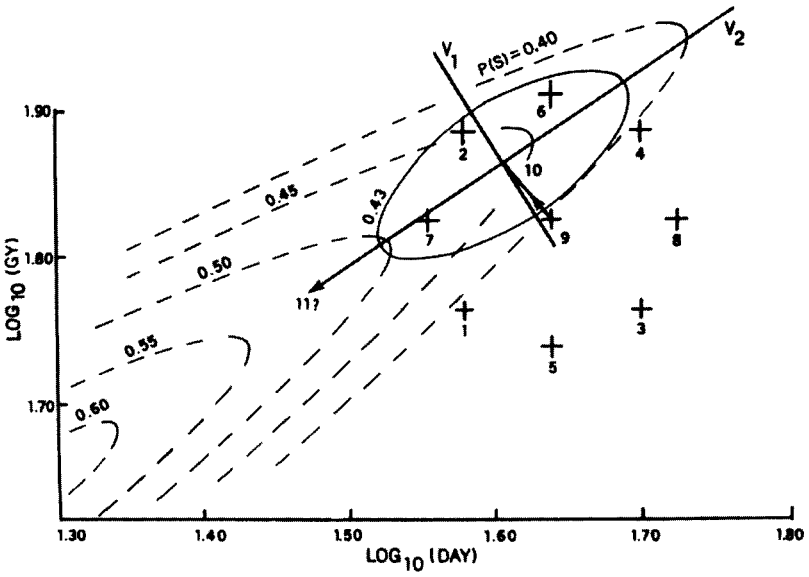

FIG. 3. A super-position of the quadratic form constructed on dose–response data obtained in a response surface design (points 1–9) in the region of D–T treatment space in which the response, $P(S)$, varies with treatment, (D,T), along a ''rising ridge.''

expensive in terms of time, effort, and money, and in the case of rapidly changing technology such as IMRT, their inherent low levels of external validity are further reduced by the passage of time so that the results at the end of the trial are often irrelevant to any current practice. They are also fraught with serious ethical problems. However, databases have, in principle, inherently high levels of external validity and raise few ethical issues.[37] Moreover, major advances in technology have provided the ability not only to create huge data stores (e.g., terabytes of data) but also the ability to extract high levels of information therefrom for prediction, for decision, and for understanding. This latter ability, loosely called ''*data-mining*,'' or, more precisely, knowledge discovery in databases (KDD) is a new field of statistics directed to the discovery of useful and often unsuspected patterns hidden deep within the data base and embedded in high-dimensional spaces.[38]

Nonlinear dynamics: Some investigators in the biomedical field have recently come to realize that, ''*The human body is an exceedingly complex mosaic of nonlinear dynamical systems. ... A detailed understanding of the dynamics of such systems must necessarily be carried out in the context of the mathematics of nonlinear systems*.'' L. Glass, 1991.[39] For example, several biological phenomena of particular importance to radiation oncology—''*jumps*,'' thresholds, oscillations, bistability and birhythmicity, emergence and self-organization—are unique to nonlinear systems. Moreover, in nonlinear systems, output is not only not simply proportional to input but may be qualitatively different from it, i.e., ''*More is different*.'' P. Anderson, 1987,[40] a potential problem in dose-escalation studies. Anderson's aphorism asserts that nonlinear phenomena are physically scale dependent, i.e., there is an inherent ''*volume effect*'' in nonlinear dynamical systems.

Models and metaphors: Useful models of empirical data on outcomes must be **statistical**; in particular, they must be generalized linear models (e.g., logit, probit, Normal, Poisson, etc.) constructed with the aid of Bayesian methods and regression diagnostics and sensitivity analyses on data obtained from well-designed experiments (especially factorial) and surveys (of databases).[15,41,42] However, useful models of theoretical constructs that inform the designs must be **dynamical**; (dynamics: the study of how the state of a system changes in time) in particular, *causal models* must be described as nonlinear dynamical systems (sets of coupled nonlinear differential equations, both ordinary and partial) and understood in terms of the **attractor-basin-bifurcation** metaphors of nonlinear dynamics.[43]

In contemplating the great promise that IMRT seems to offer to the practice of radiation oncology, it is useful to recall that, ''*Every important scientific advance that has come in looking like an answer, has turned, sooner or later—usually sooner—into a question*.'' L. Thomas, 1983.[44]

IV. DOSE–VOLUME-OUTCOME ANALYSIS: SOME NEW METHODS AND OPEN CHALLENGES

Joseph O. Deasy

Increasing the effectiveness of radiation therapy depends on improving our understanding of dose–volume factors affecting tolerance and local control. We briefly discuss some new methodologies for analyzing dose–volume effects, and some challenging open problems in this interesting field of research.

New treatment principles may emerge from a deeper understanding of dose–volume tumor and normal tissue response. What dose distributions are most desirable? For gross disease, initial theoretical studies by us, Goitein *et al.*, and Tome and Fowler, indicate that *boosting a fraction of the gross tumor volume (GTV) could significantly increase local control*.[45–48] The theoretical arguments and calculations can be summarized as indicating that: partial tumor boosts of more than 10–15 Gy would be expected to eliminate (nearly) all recurrences in which disease recurs in the boosted volume only (and not in the nonboost volume). The resulting dose response curve is expected to be nearly equivalent to that

which would hold for the same tumor but of a reduced volume equal to the nonboost gross tumor volume.[46] It is not the goal, of course, to give a reduced dose to part of the tumor, but these considerations support the use of a partial volume boost when a part of the tumor must be given reduced dose, due to constraints from nearby normal tissues. Very few clinical tests of theoretical ideas regarding TCP have been published to date,[49–51] but this is expected to change with the widespread implementation of 3-D treatment planning. However, as pointed out elsewhere in this report by Ten Haken and Yan, dosimetric ''corruption'' due to positioning errors may need to be controlled or otherwise taken into account to accurately model TCP, especially as TCP is expected to be highly sensitive to even small cold spots.

It has also been shown that the effect of hot or cold spots on TCP depends on the underlying model assumptions concerning interpatient and intratumor radiosensitivity heterogeneity.[52] In general, TCP models that neglect interpatient variations in radiosensitivity, yet still try to fit dose response curves by positing a small number of resistant clonogens, are less sensitive to cold spots. A further complicating factor is the likely existence of variations in the intratumor radiosensitivity heterogeneity distribution between patients. Attempts to reconcile mechanistic TCP models with measured *in vitro* surviving fractions at 2 Gy (SF_2), using tumor biopsy material, have shown a discrepancy.[52] Theoretical predictions of the steepness of local control curves were carried out using a TCP model based on Poisson cell kill and laboratory measured SF_2 coefficients of variation (typically 40%). Over a wide range of possible parameter values, including wide variations in mean tumor radiosensitivity, the TCP model using *in vitro* SF_2 coefficient of variation data (i.e., coefficients of variation of 35%–45%) predicted far shallower local control curves than those clinically observed. The inclusion of any other type of heterogeneity in the TCP model, such as clonogen number interpatient heterogeneity, would only make predicted dose-response curves more shallow and therefore worsen the disagreement. The only suggestion thus far to reconcile this fundamental disagreement is that the number and radiosensitivities of the more radiosensitive clonogens within a tumor (those that tend to affect SF_2 the most) varies more between patients than the number and radiosensitivities of the most resistant clonogens (those that are more likely to affect the clinical outcome).[52] Partial support for this idea is supplied by the results of Britten *et al.*,[53] who measured the intratumor heterogeneity distribution for three cervical uterine carcinoma patients, and observed intratumor heterogeneity distributions similar to those hypothesized in Ref. 52. More measurements of intratumor heterogeneity for a greater number of patients would be needed to resolve this conflict. Such investigations are needed to place models of TCP, and consequently their clinical predictions, on a more reliable footing. Another important source of tests for TCP models could potentially be small animal experiments, using nonuniform but well-characterized doses of radiation to explanted xenografts.

The biological assumptions that should go into mathematical predictions of normal tissue complication probabilities (NTCPs) are far more tenuous than those for TCP, and therefore we expect to rely mainly on clinical data and simple descriptive models with few *a priori* assumptions.

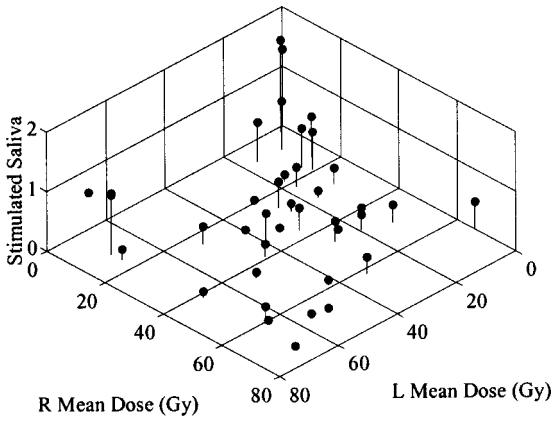

FIG. 4. Relative stimulated saliva measured six months after the end of radiation therapy, plotted as a function of mean doses to the left and right parotid salivary glands.

The present uncertainties in model assumptions and parameter values for dose–volume models are, of course, a significant barrier to clinical use of the models. This is mainly the result of the lack of high-quality clinical data for which the dosimetry is accurate and the outcome data is available, but there are also important open issues with respect to data analysis. Once a ''good'' dataset has been established for a given toxicity end point, it will, as usual, be important to understand the relative uncertainty of model fits to measured data. One very flexible method for estimating model parameter uncertainties is the bootstrap technique.[54] This method is based on the idea that a multidimensional histogram of the actual set of patient data, including outcomes, can be used as an approximation of the probability of collecting similar data in a future trial. One can then draw random samples from the present collection of data to form pseudoclinical datasets. A pseudoclinical dataset will typically include multiple copies of the data from one or more patients, by chance. The fitting process is then applied to each pseudoclinical dataset to derive the fitted parameters. This sampling–fitting process is repeated many times (typically several hundred to a thousand times) and results in an approximate probability distribution of the fitted parameter values (assuming, of course, that the model is capable of fitting, or describing, the data), given the observed input data. As an example, we have applied the bootstrap technique to analyze the reduction of saliva due to parotid gland irradiation.[55] Figure 4 shows saliva data (stimulated by chewing) at six months post-radiation therapy, plotted as a function of the left and right parotid mean doses. For didactic purposes we consider the simple model (other models ultimately fit the data better):

$$f = \frac{1}{2N_R}\sum_{i=1}^{N_R} \exp(-ad_{R,i}) + \frac{1}{2N_L}\sum_{i=1}^{N_L} \exp(-ad_{L,i}), \quad (4)$$

where f is the predicted saliva level after treatment relative to

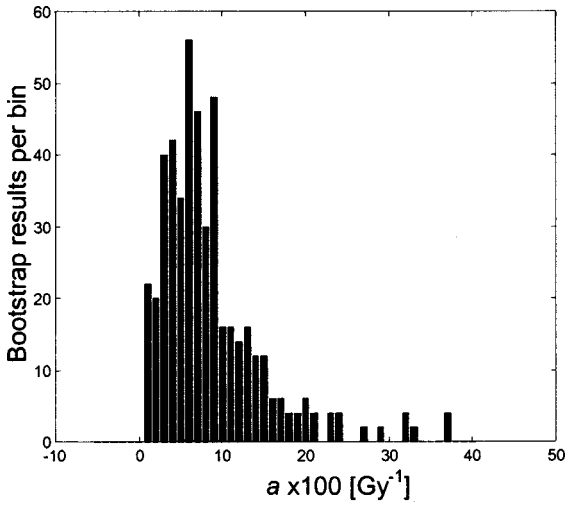

FIG. 5. Shows a histogram of 500 bootstrap resampling determinations of the fitted parameter a for Eq. (4). The bootstrap results are instructive, especially when distributions are non-Gaussian, as is the case here. Non-Gaussian distributions are common when a parameter value has a natural lower limit (such as zero) and the coefficient of variation is of the same order of magnitude or larger than the mean value. The bootstrap histogram can be used to determine confidence limits. Further analyses included a quadratic dose term (see Ref. 9).

pretreatment, a is a radiosensitivity parameter, N refers to the number of equivolume voxels in the left and right parotids, and the d_i are the voxel doses with the additional subscript indicating a left or right gland. This is a very simple FSU-type model, because the basic assumption is that functional contributions are independent from voxel to voxel. The only fitting parameter is a. Figure 5 shows the bootstrap results and thereby indicates how a could reasonably be expected to vary for other samples of the patient population.

It will be important to propagate the uncertainties inherent in the model parameter estimates into the final estimates of TCP/NTCP or another predicted outcome (such as the saliva level), so that error bars are available as part of the treatment planning process. We have developed a technique for incorporating data uncertainties into outcome estimates, which also uses the bootstrap method.[56] Instead of computing TCP, NTCP, or some other outcome measure using just the maximum likelihood fitted parameters, we compute outcome estimates for the entire list of bootstrap determined sets of parameters. Again, this list is considered to be a surrogate for the probability distribution that a given parameter set is correct. Loss of saliva function (say) would be computed for each of the bootstrap parameter determinations. The list of predicted saliva levels could be used to put error bars around the estimate of predicted saliva level.

For outcomes that are defined as probabilities (such as TCP and NTCP), we suggest that it is actually more fundamentally correct to *average TCP/NTCP over all the results based on bootstrap resampling parameters*, or some other method of estimating parameter probabilities. This would give a more faithful estimate of the actual TCP or NTCP, because we really do not know which model parameters are "correct." The minimum of the likelihood surface for the original dataset is typically broad, and it follows that model parameters that significantly deviate from the maximum likelihood values have some probability of being more correct in reality.

This bootstrap-uncertainty method will not, of course, address the ubiquitous problem of imperfect model assumptions. On this point it should be emphasized that comparisons of results using models of varying degrees of mechanistic sophistication and varying numbers of parameters is important. As the quality of the data improves, the sophistication of the models may need to be increased. This could involve more sophisticated assumptions and/or an increased number of parameters. For the parotid saliva analysis, a wide range of models (more than 20) were examined, and it was found that, with the data at hand, at least three different two-parameter models described the data to very similar accuracies.[55,57] It is desirable that a range of fitted models be reported, to give readers and other investigators a clearer idea of what sort of models describe the data well and, almost equally important, what sort of models fit the data poorly. Useful metrics for model selection include metrics based only on the relative ranking of treatment plans, such as Spearman's rank correlation coefficient[58] (in cases where the severity of the outcome is graded), and the "adjusted mean square error" (AMSE), which attempts to estimate the prediction error the model will have for new datasets. AMSE is equal to (residual square error)/$(n-2p)$, where n is the number of data points and p is the number of model parameters. The factor $1/(n-2p)$ is a more severe penalty for an increased number of model parameters than the unbiased variance factor $1/(n-p)$. Other methods for estimating model prediction error can be found in Ref. 54. Levegrun *et al.* have used receiver operating characteristic curves to judge the goodness of various predictive models.[51]

An important related, but unaddressed, problem is that the models can naively be applied to any dose distribution, whereas the data typically have been (and will probably be) derived using dose distributions that are clustered (similar) in terms of their geometrical characteristics. Inevitably, the models will be applied in cases where the dose distributions are very different from those that comprised the clinical dataset upon which the model and parameter selection was based. This might be termed the "terra incognito" problem. We will need new statistical methods to recognize when the plan evaluation is highly extrapolative as opposed to more safely interpolative. In general, it appears possible that we could engineer much stronger links between the original clinical data and the plan ranking process, and this is an attractive area of potential investigation.

Although tolerance models (i.e., NTCP) have typically been used to analyze complication data, some types of data, such as saliva reduction due to parotid gland irradiation, are probably best described by bioeffect models that describe how some biological end point varies with dose–volume changes.[55–57] For bioeffect models, data at doses much less than the typical tolerance doses can be used to understand the trend toward unacceptable toxicity. Any measurable quantity that continuously changes with dose and which is

thought to be related to an important outcome might be an interesting candidate for investigation. Functional imaging may also be used to determine other useful bioeffect metrics (such as lung perfusion[59]). Last, we emphasize that if a relevant bioeffect marker can be developed for a given outcome, the statistical precision of fitted parameters, and the sophistication of the models that can be investigated, improve dramatically. This is due to the fact that all patients yield informative data, not only on the presence of radiation damage, but also on the magnitude of the damage.

Time and fractionation effects will continue to play an important role. One underinvestigated yet potentially important effect is cellular repair on a time scale of minutes rather than hours. DNA double-strand break repair cannot be described as the result of a single exponential time constant.[60] Cannay and Millar conclude that there are roughly two components to cellular repair: one fast (with a half-time of the order of 10–15 min) and one slower (with a half-time the order of several hours).[61] They estimate that if such nonmonoexponential repaire kinetics occur in human tissue systems the probability of acute complications could be affected by approximately 25%. Recent calculations show that the difference in biologically effective dose for late-responding normal tissues between delivering 2 Gy to any tissue volume in 1 min versus spreading out or pulsing delivery to the same tissue volume over, say, 20 min could be on the order of 5%–10%.[62] Clearly dose-rate effects are an open issue for IMRT and stereotactic treatments whose session delivery times vary significantly. Potential dose-rate effects are yet another reason why IMRT delivery efficiency should be improved so that delivery times are consistent with non-IMRT delivery durations.

The evolution of dose–volume-outcome models toward becoming useful clinical tools would be aided greatly if researchers were to archive, in a publicly available fashion, all of the relevant treatment planning and outcome data upon which publications are based. This would allow workers to compare new results (datasets and models) with old, and would thereby be highly useful for further improving the scientific basis for dosimetric treatment planning decisions.

V. THE EFFECT OF PATIENT/ORGAN POSITIONING VARIATION ON TREATMENT AND PLANNING EVALUATION

Di Yan

Treatment variation in positioning of the patient/organ with respect to the treatment beams causes a temporal dose variation in critical normal tissues adjacent to the treatment target. Consequently, this temporal variation induces uncertainties in understanding the normal tissue dose response, thereby preventing reliable treatment evaluation and optimization. Numerous studies have been performed in the last decade to evaluate the potential effect of the temporal dose variation. However, the results are quite inconclusive and controversial. Besides patient-related factors, disagreements have been caused by the models and assumptions applied in each study. The following discussion outlines the quantification and evaluation models for studying the temporal dose variation, and their potential limitations.

Quantification: The geometric configuration of an organ of interest in pretreatment planning is determined by using positions of its subvolumes at the time of treatment simulation. Internal organ motion/deformation and patient setup error during the process of radiation delivery result in the displacements of these subvolumes with respect to their planning positions. Therefore, treatment variation in positioning of any organ can be quantified using the displacements of its subvolumes from the planning positions to the actual treatment positions.

The displacements can be determined based on the nature of the geometric variation in question, using a rigid body motion model and/or a nonrigid body motion model. Rigid body motion has a distance conserving property for any pair of subvolumes, and therefore implicitly assumes that patient organs maintain their shape and size during the treatment course. The rigid body motion model has been applied to measure the relative position of patient rigid bony structure with respect to the radiation beam in computing patient setup error.[63] In this case, a linear transformation, determined using the displacement of a few fiducial points in the bony structure, can be used to represent the entire structure's motion. However, the assumption of rigid body motion is limiting and unnatural for a human soft organ.

Using nonrigid body motion to study patient soft organ motion/deformation is a more appropriate model and has received more attention recently.[64–66] Unlike the rigid body motion model, the nonrigid body motion model groups all subvolumes in a deforming organ under a biomechanical structure with tissue elasticity and/or compressibility. Each subvolume displacement is then calculated by applying the finite element method.

Characterization: In the radiation treatment process, the patient organ geometry as manifested on treatment simulation image is used in pretreatment planning as the reference for treatment delivery. Any deviation between the actual treatment positions of an organ subvolume and its reference position represents an unfavorable displacement caused by either internal organ motion/deformation or patient setup error at the treatment. For a given patient, the difference between the mean of treatment positions of a subvolume and its simulation position has been used to quantify the systematic displacement $\boldsymbol{\mu}$ of the subvolume position. Similarly, each treatment position subtracted by the mean has been defined as the random displacement of the subvolume. The systematic displacement and the standard deviation $\boldsymbol{\sigma}$ of the random displacement have been commonly used to characterize a patient-specific geometric variation, such as daily setup error of the individual.[67]

General organ displacement with multiple patients' data can be characterized using a preselected fiducial point common to each patient. The fiducial point has typically been selected to represent either the average motion or the maximum motion of the organ in a given direction. When organ motion data for a large number of patients is pooled, one can

TABLE I. Treatment setup error (mm) for 30 prostate patients.

	Group 1		Group 2		All	
(mm)	Syst $\sigma(\mu)$	Rand $\sigma(\xi)$	Syst $\sigma(\mu)$	Rand $\sigma(\xi)$	Syst $\sigma(\mu)$	Rand $\sigma(\xi)$
AP	2.0	1.5	4.0	3.0	2.9	2.2
SI	1.6	1.2	2.2	2.2	1.9	1.8
RL	1.5	1.5	2.4	2.8	1.8	2.2

classify the patients into separate groups based on the magnitude of the standard deviation of fiducial point displacement. The sequence of position displacements during the treatment for the patients in a given group can be mathematically modeled as a random process, in which each patient has different systematic displacement based on his or her simulation position, and a similar standard deviation of the random displacement. With this classification, one can observe and identify an important property for intertreatment geometric variation: *the systematic displacement among patients in a specific group has a probability distribution similar to the distribution of the random displacements, if similar procedures for patient setup and organ location are applied for both simulation and treatment*. This property was observed and discussed by Bijhold in the early 1990s[67] for a study of patient setup error, and it is also true for the general displacement of an organ subvolume during the treatment course.[68] Tables I and II show a specific example of intertreatment prostatic motion obtained from a previous study.[69] This property also indicates that a group of patients who have the broadest distribution of the random displacement in their daily setup position and/or internal organ location will also have the broadest distribution of their systematic displacement. Therefore, treatment planning evaluation and generic CTV-to-PTV margin design should consider this feature.

Dosimetric effect: Dose deviation in organs of interest caused by patient/organ geometric variation can be decomposed into two parts and considered independently.[69] The first part represents the deviation in the absorbed dose at a spatial point due to patient configuration (e.g., skin surface) and/or internal tissue density distribution changes (e.g., density of the lung and the other hollow organs). These changes result in a discrepancy between the dose calculated at treatment planning and the dose actually delivered at the same spatial point. This effect could be significant for a hollow organ or organ adjacent to bone, and can be better determined by reconstructing dose based on an on-line treatment CT image. Here, it is assumed that the variation of machine's output and errors of dose calculation are minimal, and negligible. The second part represents the dose deviation due to the actual position displacement of the subvolume with respect to the anatomical landmarks, which can be determined using the knowledge of subvolume displacement calculated by considering organ motion/deformation and setup error. In the following discussion, we will focus on the dose deviation due to subvolume displacement alone.

TABLE II. Target (prostate+sv) COM motion (mm) for 30 prostate patients. COM=center of mass.

	Group 1		Group 2		All	
(mm)	Syst $\sigma(\mu)$	Rand $\sigma(\xi)$	Syst $\sigma(\mu)$	Rand $\sigma(\xi)$	Syst $\sigma(\mu)$	Rand $\sigma(\xi)$
AP	3.3	1.9	4.6	3.3	3.6	2.6
SI	2.3	1.7	3.3	2.4	2.5	2.0
RL	NA	NA	NA	NA	NA	NA

The cumulative dose deviation in a subvolume can be calculated by directly convolving the dose distribution with the distribution of organ subvolume displacement,[70] or approximated[69] as

$$\Delta D \approx \boldsymbol{\mu}^T \cdot n \cdot \overline{\nabla} d[\mathbf{x}_0, \mathbf{x}_0 + \boldsymbol{\mu}] + \boldsymbol{\sigma}^T \cdot \frac{(n-1)}{2} \cdot \frac{\partial^2 d}{\partial \mathbf{x}^2}(\mathbf{x}_0 + \boldsymbol{\mu}) \cdot \boldsymbol{\sigma}, \tag{5}$$

where $\mathbf{x}_0$ represents the subvolume position at the treatment simulation, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the systematic displacement and the standard deviation of the random displacement, n is the number of treatment fractions, and $\overline{\nabla} d[\mathbf{x}_0, \mathbf{x}_0 + \boldsymbol{\mu}]$ and $(\partial^2 d/\partial \mathbf{x}^2)(\mathbf{x}_0 + \boldsymbol{\mu})$ are the mean dose gradient in the interval $[\mathbf{x}_0, \mathbf{x}_0 + \boldsymbol{\mu}]$ and curvature of the dose distribution at the point $\mathbf{x}_0 + \boldsymbol{\mu}$. Using this relation, the cumulative dose deviation can be evaluated approximately by using the systematic displacement and the standard deviation of the random displacement alone, without considering the specific distribution of the displacement. Therefore, we can conclude the following: *the cumulative dose deviation for a organ subvolume is more sensitive to the shape of the dose profile (relative to the dose gradient and curvature), and less sensitive to the shape of the subvolume displacement distributions, as long as those distributions have equal systematic displacement and the same standard deviation of the random displacement*.

The cumulative dose deviation caused by the geometric variation, or the resulting deviation of the normal tissue complication probability (NTCP), has been commonly applied to evaluate normal tissue dose responses and treatment plans. However, this evaluation potentially masks the effect of the random displacement, particularly for critical normal organs adjacent to the treatment volume. This is because the adjacent normal organs are commonly located at the linear portion of the dose penumbra region, where the dose distribution curvature is small, but the dose gradient has its maximum. Figure 6 indicates the frequency distribution of fraction dose deviations, as well as the corresponding cumulative dose deviations, to two subvolumes, which are located in the treatment planning on the 40% and 90% isodose surface, respectively. The fractional dose deviations were calculated by simulating the subvolume displacement as a Gaussian distribution with zero mean, representing no systematic displacement, and 5 nm standard deviation, characterizing the random displacement. The result demonstrates that the spectrum of fraction dose deviations is broad, with more than

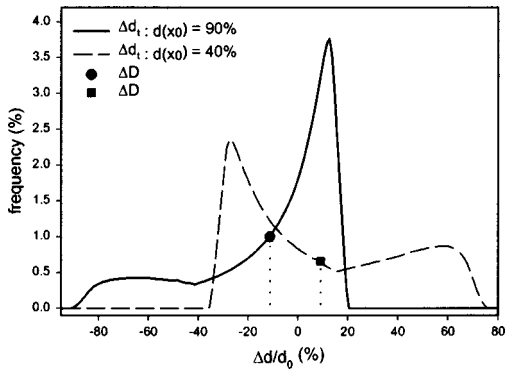

FIG. 6. Cumulative dose deviation ΔD and the distribution of fraction dose deviation Δd_t in a subvolume. Both deviations are normalized to the corresponding planned dose.

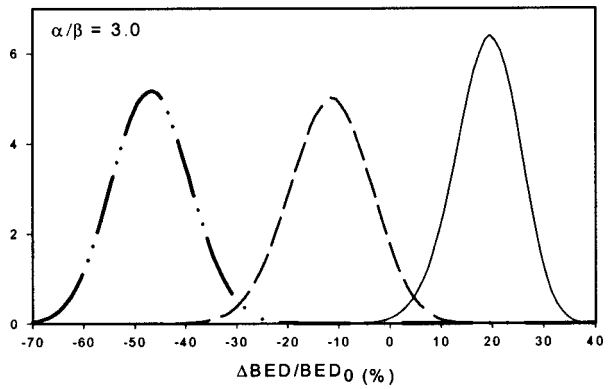

FIG. 7. Distribution of BED deviations in a subvolume for three groups of patients. The displacement in each group of patients has μ and σ to be (4 mm, 5 mm) (dot-*n*–dashed curve), (0 mm, 5 mm) (dashed curve) or (−4 mm, 5 mm) (solid curve), respectively.

100% difference between the maximum and minimum values, however the corresponding cumulative dose deviation is less than 10%. Furthermore, a similar or even identical cumulative dose deviation can arise from totally different spectrums of fraction dose deviations. This implies that using the cumulative dose deviation alone to perform treatment evaluation ignores the effect of fraction dose deviation, and could misinterpret the treatment dose response for organs of interest, particularly for critical normal organs. Moreover, this issue becomes even more problematic when a treatment has a small targeting margin and sharp dose gradient.

Radiobiological effect: Effect of the fraction dose deviation in an organ of interest can be evaluated using the biological effective dose (BED),[71,72] which is the equivalent total dose if given in infinitely small fractions. The deviation of BED for an organ subvolume calculated using the planned dose and delivered dose can be denoted[69] as

$$\Delta\mathrm{BED}=\left(1+\frac{2\cdot d_p}{\alpha/\beta}\right)\cdot\Delta D+\frac{1}{\alpha/\beta}\cdot\sum_{t=1}^{n}[\Delta d_t]^2, \tag{6}$$

where d_p is the dose in the subvolume initially calculated at treatment planning, therefore the corresponding BED at the treatment planning is $\mathrm{BED}=n\cdot d_p\cdot(1+d_p/\alpha/\beta)$, ΔD is the cumulative dose deviation, and Δd_t is the fraction dose deviation to the subvolume between the actual delivery at the treatment t and the planning calculation. Therefore, the radiobiological dose deviation ΔBED is proportional to both the cumulative dose deviation and the square of the fraction dose deviation. Furthermore, the BED deviation is more pronounced for late reacting normal tissues, due to the smaller α/β values. Therefore, *significant deviation on the biological effective dose could be expected in a critical normal structure, even if the cumulative dose deviation in this structure is negligible. Patients with similar geometric variation characteristics can experience different BEDs, and the differences are very sensitive to the dose gradient and the total number of treatment fractions*.

Figure 7 shows the distribution of BED deviations among patients to a subvolume located at the 80% isodose line during the treatment planning. The value in Fig. 7 was calculated for three groups of patients with the subvolume displacement (in one dimension) of each group: $(\mu,\sigma)=(4,5)$, (0, 5), or (−4, 5) mm, respectively. The frequency of patients who experience different BEDs was plotted by normalizing their ΔBEDs to the planned value, with the prescribed fraction dose 2 Gy for total treatment fractions $n=35$. The results demonstrate that a broad spectrum of biological dose response variation can occur in clinical treatment, even if a given group of patients have similar geometric variation characteristics.

VI. EVIDENCE OF VOLUME EFFECTS FOR RECTAL BLEEDING

Andrew Jackson

Treatment planning for the external beam radiotherapy of cancers rests on a few simple principles. One of the most important is that tolerance doses for normal tissues depend on the exposed volumes of the involved organs (i.e., there are volume effects). This assumption provides the basis for the many Phase I dose escalation trials currently in progress, which use the information provided by CT scans and 3-D dose distributions to design treatments that minimize normal organ exposure while raising the dose to the target. The existence of volume effects is widely held to be true, but for clinical endpoints the supporting evidence is often of poor quality and sparse. There are clear reasons for this; some historical and some intrinsic to the nature of complications.

A specific example will illustrate these problems: evidence for volume effects in rectal bleeding after radiotherapy. When severe, rectal bleeding can be a dose limiting late complication of external beam radiotherapy for prostate cancer.

Estimates of partial volume tolerance doses were compiled by Emami *et al.*[73] and gathered mostly from experience of complications that occurred before the era of 3-D treatment planning, when little or no quantitative data was available on the exposed volumes of involved organs. At that time, Emami *et al.* stated that there was no volume effect for severe late effects in the rectum, while admitting that published articles contained little or no volume information. Ma-

meghan *et al.*,[74] who studied 218 patients treated for prostate cancer to various doses in the range 50–65 Gy, saw a correlation between bowel complications and inclusion of the whole pelvis in the irradiated volume. Dearnaley *et al.*[75] conducted a randomized clinical trial of nonconformal and conformal radiotherapy for prostate cancer and demonstrated a decrease in late rectal bleeding (RTOG grade 2 or higher) from 15% to 5%.

Dose–volume data from 3-D conformal dose escalation studies have also indicated that volume effects for rectal bleeding exist. Benk *et al.*[76] reported data from Dose Volume Histograms (DVHs) of 41 patients treated by photon whole pelvis and proton boost fields to 75.6 Cobalt Gy Equivalent (CGE), with the rectum localized using a probe. Fourteen patients developed rectal bleeding, (6 grade 1, and 8 grade 2 using the RTOG classification[77]). A significantly higher incidence of bleeding was observed among those patients with $\geqslant$40% of the anterior rectal wall receiving 75 CGE. Subsequent analysis of the same data by Hartford *et al.*[78] showed that dose–volume cutoffs at lower doses (down to $\geqslant$70% of the anterior wall receiving 60 Gy) also produced significant correlation with bleeding, as did a model of NTCP. Schultheiss *et al.*[79] reported high actuarial rates of grade 2 and 3 GI morbidity following prostate treatment using a four-field technique, showing a strong dose response. For patients treated to 73–76 Gy to the isocenter, a significant reduction in morbidity was observed in those patients for whom rectal shielding was increased for the last 10 Gy of treatment. It was not clear if this was due to a volume effect, or to the decrease in dose to the rectal wall, since the DVHs were not analyzed. Boersma *et al.*[80] studied DVHs from 130 patients treated for prostate cancer to isocenter doses between 70–78 Gy in 2 Gy fractions with three-field 3DCRT. Pelvic nodes were treated to 64% of the prescription dose at a lower dose per fraction using a simultaneous boost technique delivered with partial transmission shielding, and doses above 70 Gy were delivered with additional rectal shielding. Grade 2 or higher rectal bleeding (RTOG) was observed in 18 patients, four of whom required one or more laser treatments and blood transfusions. No significant correlation with dose–volume parameters was seen for grade 2 rectal bleeding. However, for the four cases of severe rectal bleeding, a significant correlation with volume receiving doses greater than 65 Gy was seen. Recently, Skwarchuk *et al.*[81] and Jackson *et al.*[82] have studied patients treated prone with a six-field technique to 75.6 Gy at Memorial Sloan–Kettering Cancer Center. These studies included 36 patients with grade 2 rectal bleeding by 30 months and a random sample of 83 out of 192 eligible patients without grade 2 bleeding at 30 months. The percent volumes of rectal wall exposed to 47 and 77 Gy were both found to be significantly correlated with grade 2 rectal bleeding. Additional analysis of this data indicates that a functional reserve of tissue receiving less than 40–50 Gy may be important in preventing rectal bleeding.[83] A recent article by Fenwick,[84] which found a dependence of rectal bleeding on the area of rectal wall receiving $\geqslant$57 Gy lends support to the hypothesis that the extent of exposure to relatively low doses may play a role in causing rectal bleeding.

Properly conservative treatment practice naturally limits the number of serious complications (''responders'') that arise, severely limiting the statistical power of studies from single institutions. While studies without responders may help to define regions of safe treatment, only studies with responders can locate the boundary of safe treatment. In all the recent reports, the raw numbers of severe cases of rectal bleeding reported were small: Benk/Hartford *et al.*: 1 case of grade 3 bleeding; Boersma *et al.*: 4 cases requiring laser surgery/transfusion; Schultheiss *et al.*: 15 cases of grade 3 morbidity. Studying the volume effect in such small numbers of patients all treated with a similar technique is very difficult. Studying grade 2 complications has the advantage of better statistics. However, the grade 2 endpoint is not dose limiting, and its diagnosis is more subjective. Some justification for studying this end point comes from the study of Schultheiss *et al.*, which showed that the rate of grade 3 GI complications also rose when the grade 2 rate rose above 20%.

Data from single institutions, often involving single treatment techniques, may not provide an adequate range of dose volume combinations from which to determine volume effects. For example, while Schultheiss *et al.* found that treatment of the whole pelvis to 45 Gy was not significantly associated with increased GI morbidity (in contrast to the results of Skwarchuk *et al.* and Jackson *et al.*), the studies of Benk *et al.*, Hartford *et al.*, and Boersma *et al.* could shed no light on this issue, since all patients received irradiation with large pelvic fields to 44–50 Gy.

Other (perhaps intractable) problems arise when comparing results from different institutions. For example, organ motion may be expected to change the dose delivered from the planned dose, but the extent and location of these effects may depend upon institution-specific factors such as patient setup (e.g., supine or prone) and treatment technique (e.g., patients in the studies of Benk *et al.* and Hartford *et al.* were treated with a perineal proton boost field with the rectal wall fixed with a probe). Finally, and importantly, endpoints used to study the outcome are difficult to define (especially for low-grade complications) and may differ not only from institution to institution, but also from physician to physician.

Despite the limitations imposed by poor statistics, treatment technique, organ motion, and the difficulties inherent in comparing results from different institutions, evidence is emerging, from 3-D conformal dose-escalation trials, that demonstrates the existence of volume effects in late rectal bleeding. Continued accumulation and careful analysis of this data is the only way we might hope to quantify these important limitations on external beam radiotherapy for prostate cancer.

VII. USING DOSE–VOLUME MODELS TO DESIGN CLINICAL TRIALS: NORMAL TISSUE-BASED DOSE ESCALATION IN LIVER AND LUNG

Randall K. Ten Haken

Liver and lung belong to a group of normal tissues generally believed to (a) exhibit a ''volume effect,'' (b) sustain

damage in a ''parallel'' versus ''serial'' fashion, and (c) be subject to a ''functional reserve'' below which damage may not cause injury and above which injury results. This makes their study attractive for NTCP modeling. Liver studies represent perhaps the best hope for modeling. The volume and dose can be estimated fairly well, local damage may exhibit a response to dose, and the relationship of cumulative local damage to organ injury may also be indicated. Due in part to a desire for aggressive treatment, the consequences of lung irradiation have received more attention. However, modeling will remain challenging due to many issues associated with the assessment of lung complication data. Issues include the definition of an end point (pneumonitis, fibrosis, etc.), confounding factors such as preexisting disease and chemotherapy, modeling lung function (nonuniform distribution of FSUs, regional differences in FSU radiosensitivities and function), and compensatory effects. Functional imaging[85–88] can help in the assessment of some of these effects. However, dose distributions in both liver and lung are suspect due to organ motion due to patient breathing.[89] Despite these concerns, general approaches to systematically gather patient data are needed and are starting to emerge. Thus, treatment planning and better understanding of normal tissue complications remain as challenging issues for those involved in the treatment of lung and liver cancer,[90–100] to list a few.

Variations on one general approach toward implementing NTCP models for liver and lung have been adopted at several treatment centers.[90,94,98,101,102] First, one treats patients and collects 3-D dose–volume data, together with the assessments of patient outcome. Next, a retrospective analysis can be performed to estimate parameters of a descriptive NTCP model.[103,104] Finally, a prospective ''normal tissue dose escalation'' trial may be started, escalating groups of patients from nominal iso-NTCP levels according to common levels of risk. This contrasts with standard dose trials that deliver a target dose without regard to the volume of normal tissue. As a consequence, these types of trials accommodate the introduction of new technologies that may produce more conformal dose distributions. Although the target volume dose distribution may change, corresponding changes in the prescription dose ensure that the same normal tissue risk levels are maintained. Thus, the data from a new patient, treated with an advanced technology, is guided by the same iso-NTCP level as the previous patient treated with the older technology. As more outcome data are acquired, it should then be possible to continue to refine model parameters or use the data for input to test other models.

Recent trials in the treatment of intrahepatic tumors help to illustrate this approach. Substantial retrospective data suggested that parts of the liver could safely receive far higher doses than traditionally delivered. In 1987 a series of studies using 3-D conformal therapy were begun based on two fundamental concepts. First, the ability existed to significantly reduce the dose to the normal liver.[105] Second, conformal treatment planning permitted quantification of the fraction of normal liver irradiated that could be conveniently expressed for input in a NTCP model. The first trial used a simple scheme based on the volume of a normal liver receiving $>50\%$ of the prescription dose.[106] For both colorectal cancers metastatic to the liver and primary hepatobiliary cancers, far higher tumor response rates were observed than had previously been possible using radiation.[107] Retrospective analysis of those data allowed the refinement of parameters of the Lyman model[108] to describe the probability of causing radiation-induced liver disease, based on both the radiation dose and the liver volume irradiated.[109] Subsequently, a new prospective trial was initiated in an attempt to safely escalate the dose of radiation for patients with intrahepatic cancer. The trial uses a protocol[98] in which each patient receives the maximum possible dose while being subjected to a preset risk of radiation-induced liver disease (RILD) or ''radiation hepatitis'' based on a normal tissue complication probability (NTCP) model.[108,110] The goal of planning is to maximize the dose to the target while both minimizing the effective volume Veff for the liver and respecting other dose-limiting organs. This is made possible through the realization[102] that the computation of Veff is independent of dose ''units.'' That is, the value of Veff depends only on the shape of the DVH and the relative value of Dref. Therefore, a value of Veff may be computed for each patient from a relative isodose distribution (%) before a physical dose (Gy) (based on an iso-NTCP level) is prescribed. Early results indicate[91,92,98] the dose delivered using this approach is significantly higher than the dose that would have been delivered by the previous protocol, and data for use in NTCP analysis continues to be accrued. These results suggest that a NTCP model based on patient data (rather than literature estimates) can be used prospectively to safely deliver far higher doses of radiation with a more consistent risk of complication than would have previously been considered possible for patients with intrahepatic cancer. However, clearly, multi-institutional studies will be required to obtain significant numbers of events to permit a good parameterization of model parameters.

A major potential impediment to effectively modeling normal tissue dose-volume relationships in the abdomen and pelvis is including the effects of patient breathing in the resulting dose distributions. Treatment planning CT scans (even if done with a breath-hold) illustrate the shape and position of thoracic and abdominal structures at one point in time and thus may poorly predict the actual patient status for dose calculations unless similar breath-hold procedures are used at treatment.[111,112] The dosimetric corrections introduced by geometric changes can be as large as the uncertainties often associated with the dose calculation density correction algorithms.[113,114] Current treatment plans include a margin for ventilation, typically based on the fluoroscopic observation of the tumor shadow and/or diaphragm under ventilation at simulation. While this may ensure tumor volume coverage, it leads to larger than necessary treatment volumes and can produce misleading NTCP estimates. That is, even given that (when done properly) standard practice (PTV formation) could help assure that CTVs would indeed receive their desired doses, traditional initial dose distribution calculations would not include the effects of uncertainties from daily setup variations and organ motion. Thus, they would not represent dose distributions actually received by

patients over their course of treatment. Preliminary studies indicate that clinically meaningful (several fraction) differences in the prescribed dose could result if dose distributions that include the effects of organ motion were used in place of the original static treatment plans.[115,116] Clearly, further studies of the impact of patient-related geometric variations on normal tissue dose distributions are called for.

VIII. THE EFFECT ON PLAN EVALUATION OF UNCERTAINTY IN TOLERANCE LIMITS

Mark Langer

The levels of radiation dose that can be tolerated by different tissues are not precisely known, and this uncertainty makes it difficult to compare treatment plans. Comparisons become uncertain when a small change in a tolerance limit produces a large change in the dose delivered to the tumor. Not all tolerance limits have the same effect on tumor dose. The selection of treatment plans can be improved by focusing efforts to reduce tolerance uncertainty on those limits for which small errors cause the greatest shortfalls in tumor dose. Modern methods for plan construction allow the sensitivities of tumor dose to errors in tolerance limits to be discovered.

A wide range of tolerance limits for different organs can be found in the literature or in protocol rules. Dose limits for the spinal cord found among protocols that outline the radiotherapy to be used for a single condition—small cell lung cancer treated with coincident platinum-based chemotherapy range from 44 Gy in 22 fractions to 50 Gy in 25 fractions, a variation of 14%.[117,118] The volume fraction of the heart that protocols require to be protected is nearly always 50%, but the dose limit on this protected fraction can range from 30–50 Gy. Both the threshold dose and the protected volume of the lung to be maintained in thoracic radiation are uncertain. A cutoff of 30 Gy for up to 70% of the total lung is specified in a small cell protocol using coincident chemotherapy, while a dose cutoff of 20 Gy for a stratified set of lung volumes has been adopted by the RTOG for their tumor dose escalation protocol 93-11.[117,119] In the literature, the threshold dose level used to relate pulmonary damage to the lung volume irradiated to beyond that level has been taken to be either 30 or 20 Gy.[120,121] Some studies do not use dose–volume constraints, but rather continuous response models, as discussed above. Small changes in a tolerance limit can produce large effects on the dose that can be delivered to the tumor, but the effect of uncertainty on tumor dose need not be the same from one limit to the next. A study evaluating the benefit of a computer controlled technique for treating lung cancer found that the gain over a conventional method was very sensitive to the volume of contralateral lung permitted to receive more than a threshold dose of 20 Gy. If the lung volume allowed more than 20 Gy was restricted to 27%, the gain by substituting a computer controlled technique for conventional therapy was 7 Gy, but further tightening the volume limit to 25% reduced the gain to only 1 Gy.[122] A larger study found that the minimum target dose that could be achieved with conventional therapy could be increased in two of six cases by more than 9% to reach a level greater than 80 Gy by expanding by 2%–3% the volume of the contralateral lung allowed to receive a threshold dose of 20 Gy.[123] While gains in tumor dose for lung cancer treatment may be very sensitive to the limit on the lung volume that is allowed to receive greater than some threshold dose, the exact setting of the threshold dose level may be less important. It has been suggested that the volume of lung receiving >20 Gy and the volume of lung receiving >30 Gy may be tightly correlated parameters, at least when traditional plans are used.[124] It may be that the minimum tumor dose is more sensitive to small changes to a volume limit than to a dose limit. In this case, a phase I study will yield the greatest gain in target dose for a given reduction in the uncertainty of a tolerance limit if it is designed to reduce the error in the maximum volume allowed a given dose and not the maximum dose allowed in a given volume.

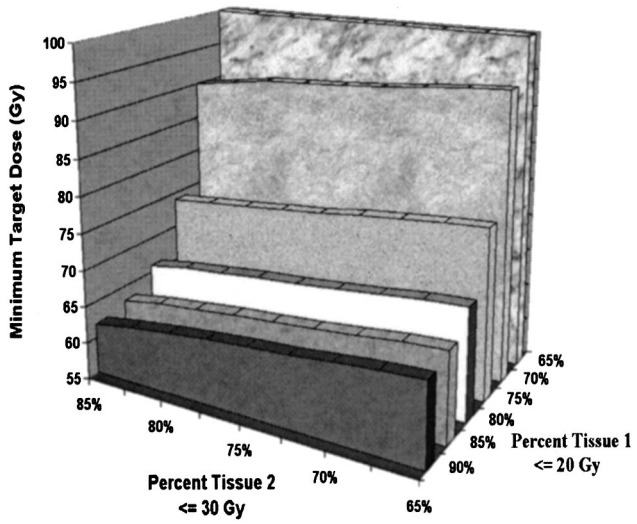

FIG. 8. Minimum tumor dose possible for different combinations of dose volume constraints on two tissues.

The sensitivity of tumor dose to tolerance uncertainty can readily be discovered using an optimization package for beam weighting that can quickly construct a treatment plan given rules on the volume distribution of dose in critical structures, including target and normal tissues. In a series of lung cancer patients whose boosts were planned using a computer controlled technique, small changes in a limit on dose inhomogeneity within the target produced large changes in the minimum tumor dose that could be delivered. A minimum tumor dose of >80 Gy could be delivered in all six cases examined when an inhomogeneity limit of 20% was accepted, but the minimum tumor dose fell to the range 44–64 Gy when the limit was tightened to 13%–17%.[125]

An example of the differences that can be seen in the sensitivity of tumor dose to various tissue limits is shown in Fig. 8 for a hypothetical model. The figure depicts the highest value for the minimum tumor dose that can be obtained under different combinations of partial volume limits for two tissues. For tissue 1, a reduction of 5%, from to 75% to 70%, in the volume held to a dose $\leqslant 20$ Gy allows a large increase

in the minimum dose that can be delivered to the tumor, about 16 Gy. On the other hand, changes in the volume of tissue 2 that must be held to a dose $\leqslant$30 Gy, within the range of 65%–85%, barely affect the minimum tumor dose that can be obtained. If similar findings were seen in clinical series, then a planned phase I dose escalation trial should be designed to reduce the uncertainty in the tolerance limit for tissue 1 rather than for tissue 2 if the aim is to allow the greatest possible increase in tumor dose for a given reduction in uncertainty. In clinical applications, similar behavior might be observed in the treatment of thoracic tumors, with tissue 1 representing lung and tissue 2 representing heart. When new treatment methods are applied, small relaxations in the dose constraints that have been traditionally set may be introduced. An example is the relaxation in inhomogeneity limits that has accompanied the use of intensity modulated radiotherapy. The relaxations confound the ability to assess the merits of the new techniques. The dose gains seen might be the consequence of the relaxed dose constraints and not the new technology *per se*.

The acute sensitivity of tumor dose to small changes in a tolerance condition implies that any comparison of treatment techniques should adhere to a rigid set of rules. Not uncommonly, one finds that a new treatment technique trades off a small deterioration in the dose distribution in one structure for a larger gain in another.[126] Even if the deterioration produced in one structure is so small as to be felt to be clinically unimportant, the newer technique cannot be said to be superior. Had the older technique been allowed the same small relaxation of the dose distribution in the first structure, it might have produced the same gain in the dose distribution in the second. Relaxations as small as 2%–3% in the lung volume receiving $\geqslant$20 Gy were found in one study to completely eliminate the gains attributed to conformal therapy over standard techniques for treating lung cancer.[123] Data describing the sensitivity of the objective to changes in the constraints over a wide range of values have been recently reported for clinical cases in the abdomen and prostate.[127,128]

The sensitivity problem is not eliminated by the introduction of score functions that rate the dose distributions in different structures rather than demand that a set of constraints be satisfied. The score functions themselves have errors whose sizes are seldom estimated. Even if score functions correctly order plans according to their probabilities of producing reactions in different tissues, the overall ranking of the plans by the probability of avoiding any tissue complication may still be wrong.[129] There is a strong connection between maximizing an objective subject to constraints, and maximizing a weighted function of the constraint terms and the objective. When the feasible space is piecewise convex, the solution to the first problem corresponds to a solving the second problem for a particular set of weights.[130] If there are errors in the constraints, then there will also be errors in any function that scores plans by weighting their deviations from the constraints.

Treatment techniques are best compared by considering the sensitivity of the delivered tumor dose to errors in the specification of the constraints. If tumor control or normal tissue complication probabilities are used to compare plans, the error in their individual terms should be determined and propagated to give an error in the overall score. Phase I studies should be designed to reduce the error in the dose or volume limit that would allow the largest increase in tumor dose were its uncertainty made smaller. This determination is necessarily a computer-based exercise given typical patient datasets. The effect on tumor dose of relaxing a constraint within its range of uncertainty should be made available to physicians involved in the selection of treatment plans.

IX. SUMMARY

Widespread availability of 3-D treatment planning systems has facilitated the analysis of outcomes based on the planned dose distributions. Although volume-effects are difficult to study due to understandably small clinical-event rates, volume-effects based on fully 3-D dose distributions have been observed for rectal, liver, and lung end points, among other sites. From the work presented here it is clear that much more clinical data directed toward evaluation of 3-D conformal and IMRT dose distributions is necessary. Points of emphasis include the following.

(i) Dose–volume outcome models with a minimal number of parameters, which can describe the observed data and rank dose distributions, for both normal tissues and tumors, will potentially play an important clinical role and deserve further development.

(ii) Pooled-data analyses, which utilize data and modeling techniques from different institutions could potentially improve both the range of validity of dose–volume outcome models and the accuracy of model parameters.

(iii) Analyses of the cumulative uncertainties in the application of radiobiological outcome models to individual patients, and the presentation of such uncertainties to clinical users, is desirable.

(iv) Patient motion effects, such as breathing motion, can significantly change the response to radiation. However, in some cases such motion can be measured and modeled, and thereby included in outcome data analysis.

(v) Volume-effect models can be effectively used as a basis for dose-escalation protocols, thereby leading to better models and ultimately better protocols.

(vi) Uncertainties in volume-effect models and dose-volume constraints used for optimized treatment planning have important clinical consequences. In particular, research should focus on those aspects of models or constraints that are more likely to impact the ability to safely deliver tumoricidal doses.

In order to harvest the benefit of advances in dose delivery techniques, we require quantitative, individualized, pretreatment estimates of risks and benefits. The present situation brings to mind words of Fermi: ''...we must be prepared for a long hard pull... .'' [131]

ACKNOWLEDGMENTS

This workshop was partially supported by the National Institutes of Health (NIH). Individual investigators were also

supported by NIH grants: No. CA50628 and No. CA21239 (AN), No. CA85181 (JOD), No. CA71785 (DY), No. CA59827 (RTH), and No. 2PO1-CA-59017 (AJ).

[a]Corresponding author and address for reprints: J. O. Deasy, Department of Radiation Oncology, Box 8224, Washington University Medical School, 510 So. Kingshighway Blvd., St. Louis, Missouri 63110. Telephone: (314) 362-1420; electronic mail: deasy@radonc.wustl.edu

[1] E. J. Hall, in *Radiobiology for the Radiologist*, 4th ed. (Lippincott, Philadelphia, PA, 1994), p. 478.
[2] H. D. Thames and J. H. Hendry, in *Fractionation in Radiotherapy* (Taylor & Francis, New York, 1987), p. 297.
[3] A. Niemierko, ''Reporting and analyzing dose distributions: a concept of equivalent uniform dose.'' Med. Phys. **24**, 103–110 (1997).
[4] A. Terahara, A. Niemierko, M. Goitein, D. Finkelstein, E. Hug, N. Liebsch, D. O'Farrell, S. Lyons, and J. Munzenrider, ''Analysis of the relationship between tumor dose inhomogeneity and local control in patients with skull base chordoma,'' Int. J. Radiat. Oncol., Biol., Phys. **45**, 351–358 (1999).
[5] Q. Wu, R. Mohan, and A. Niemierko, ''IMRT optimization based on the generalized Equivalent Uniform Dose (EUD),'' in ICCR 13, 2000, Heidelberg, Germany.
[6] H. R. Withers, J. M. Taylor, and B. Maciejewski, ''Treatment volume and tissue tolerance,'' Int. J. Radiat. Oncol., Biol., Phys. **14**, 751–759 (1988).
[7] A. Niemierko and M. Goitein, ''Modeling of normal tissue response to radiation: the critical volume model,'' Int. J. Radiat. Oncol., Biol., Phys. **25**, 135–145 (1993).
[8] A. Niemierko, ''A generalized concept of Equivalent Uniform Dose,'' Med. Phys. **26**, 1100 (1999).
[9] G. J. Kutcher, C. Burman, L. Brewster, M. Goitein, and R. Mohan, ''Histogram reduction method for calculating complication probabilities for three-dimensional treatment planning evaluations,'' Int. J. Radiat. Oncol., Biol., Phys. **21**, 137–146 (1991).
[10] J. T. Lyman, ''Complication probability as assessed from dose–volume histograms,'' Radiat. Res., Suppl. **8**, S13–S19 (1985).
[11] D. E. Herbert, ''Quality assessment and improvement of dose-response models: some effects of study weaknesses on study findings. C'est magnifique?'' *A Report of Task Group 1 of the AAPM Biological Effects Committee* (Medical Physics Publishing, Madison, WI, 1993), AAPM Report No. 43.
[12] A. Niemierko, ''A unified model of tissue response to radiation,'' Med. Phys. **26**, 1100 (1999).
[13] K. S. C. Chao, J. O. Deasy, J. Markman, J. Haynie, C. A. Perez, J. A. Purdy, and D. A. Low, ''A prospective study of salivary function sparing in patients with head and neck cancers receiving intensity-modulated or three-dimensional radiation therapy: initial results,'' Int. J. Radiat. Oncol., Biol., Phys. **49**, 907–916 (2001).
[14] G. E. P. Box, ''Robustness in the strategy of scientific model building,'' in *Robustness in Statistics*, edited by R. L. Launer and G. N. Wilkinson (Academic, New York, 1979), pp. 201–236.
[15] D. Herbert, ''Quality assessment and improvement of dose response models: Some effects of study weaknesses on study findings. C'est magnifique?,'' ''Evaluation of Models for Dose-Response in Radiation Oncology,'' of the Biological Effects Committee, A Report of Task Group 1 (Medical Physics Publishing, Madison, WI, 1993), AAPM Report No. 43.
[16] L. Cohen, in Ref. 15, pp. i–iii.
[17] J. F. Williamson, P. G. Goldschmidt, and T. Colton, ''The quality of medical literature: An analysis of validation assessments,'' in *Medical Uses of Statistics*, edited by J. C. Bailar, III and F. Mosteller, (NEJM Books. Waltham, MA, 1986), pp. 370–391.
[18] M. R. Chassin and R. W. Galvin, ''The urgent need to improve health care quality,'' J. Am. Med. Assoc. **280**, 1000–1004 (1998).
[19] M. Hewett and J. Simone, in *Ensuring Quality Cancer Care*, Institute of Medicine/National Research Council (National Academic, Washington, DC, 1999).
[20] J. E. Ware, ''The status of health assessment 1994,'' Annu. Rev. Public Health **16**, 327–354 (1995).
[21] U.S. Congress Patient Outcome Research Act, 1989.
[22] F. Mosteller (personal communication, 1999).
[23] T. Chalmers, ''Problems induced by meta-analysis,'' Stat. Med. **10**, 971–980 (1991).
[24] National Research Council, *Combining Information. Statistical Issues and Opportunities for Research*, Washington, DC, 1992.
[25] T. Chalmers, ''Meta-analysis in clinical medicine,'' Trans. Am. Clin. Climatol. Assoc. **99**, 144–150 (1987).
[26] L. Bero and D. Rennie, ''The Cochrane collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care,'' J. Am. Med. Assoc. **274**, 1935–1938 (1995).
[27] C. Naylor, ''Gray zones of clinical practice: Some limits to evidence-based medicine,'' Lancet **345**, 840–942 (1995).
[28] S. B. Thacker and R. L. Berkelman, ''Surveillance of medical technologies,'' J. Public Health Policy **7**, 363–377 (1986).
[29] S. J. Senn, ''Falsificationism and clinical trials,'' Stat. Med. **10**, 1679–1692 (1991).
[30] R. Brook, ''Using scientific information to improve quality of health care,'' in *Doing More Good Than Harm. The Evaluation of Health Care Interventions*, edited by K. Warren and F. Mosteller, Ann. N.Y. Acad. Sci. **703**, 74–85 (1993b).
[31] U.S. General Accounting Office, *Cross Design Synthesis*, GAO/OEND-92-18. Washington, DC, 1992.
[32] D. Herbert, ''Response surfaces for binary events. An empirical basis for radiation oncology,'' in *Proceedings of the 6th International Conference on the Use of Computers in Radiation Therapy*, edited by U. H. Rosenow (Gottingen University Press, Gottingen, Federal Republic of Germany, 1978), pp. 210–238.
[33] D. Herbert, ''An extreme value paradigm for the effect of size on target volume on end results in radiation oncology,'' Med. Phys. **10**, 589–604 (1983).
[34] D. Herbert, ''An empirical extreme value model of the volume effect in radiation oncology,'' in *Optimization of Cancer Radiotherapy*, Proceedings of the 2nd International Conference on Dose, Time, and Fractionation in Radiation Oncology, edited by B. Paliwal, D. Herbert, and C. Orton (American Institute of Physics, New York, NY, 1984), pp. 381–401.
[35] G. E. P. Box, ''The determination of optimum conditions,'' in *The Design and Analysis of Industrial Experiments*, edited by O. Davies (Oliver and Boyd, London, 1967).
[36] D. Herbert, ''The bivariate probit model of uncomplicated control of tumor: A heuristic exposition of the methodology,'' Int. J. Radiat. Oncol., Biol., Phys. **39**, 213–225 (1997).
[37] R. M. Califf, D. B. Pryor, and J. C. Greenfield, ''Beyond randomized clinical trials: Applying clinical experience in the treatment of patients with coronary artery disease,'' Circulation **74**, 1191–1194 (1986).
[38] D. J. Hand, ''Date mining: Statistics and more.'' Am. Stat. **52**, 112–118 (1998).
[39] L. Glass, ''Nonlinear dynamics of physiological function and control,'' Chaos **1**, 247–250 (1991).
[40] P. Anderson, ''More is different,'' Science **177**, 393–396 (1972).
[41] D. Herbert, ''Modeling the effectiveness of radiation treatment,'' in *Biomedical Uses of Radiation*, edited by W. R. Hendee (Wiley-VCH, New York, 1999), pp. 1227–1288.
[42] D. E. Herbert and D. J. Brenner, ''The use of the linear quadratic model in clinical radiation oncology can be defended on the basis of empirical evidence and theoretical argument,'' Med. Phys. **24**, 1245–1248 (1997).
[43] D. Herbert, ''Overview of nonlinear dynamical systems and complexity theory,'' in *Chaos and the Changing Nature of Science and Medicine: An Introduction*, edited by D. Herbert (American Institute of Physics, New York, NY 1996), pp. 1–34.
[44] L. Thomas, *Late Night Thoughts on Listening to Mahler's Ninth Symphony* (Viking Press, New York, 1983).
[45] M. Goitein, A. Niemierko, and P. Okunieff, ''The probability of controlling an inhomogeneously irradiated tumour: a strategem for improving tumour control through partial tumour boosting,'' presented at the 19th L H Gray Conference: Quantitative Imaging in Oncology, Newcastle, 1997.
[46] J. O. Deasy, ''Tumor control probability models for nonuniform dose distributions,'' *Proceedings of the 5th International Conference on Dose, Time and Fractionation in Radiation Oncology: Volume and Kinetics in Tumor Control and Normal Tissue Complications*, Madison (Medical Physics Publishing, Madison, 1998), pp. 65–85 (also available from http://castor.wustl.edu/~deasy).
[47] A. T. Tome and J. F. Fowler, ''Selective boosting of tumor subvolumes,'' Int. J. Radiat. Oncol., Biol., Phys. **48**, 593–599 (2000).
[48] J. O. Deasy, ''Partial tumor boosts: even more attractive than theory predicts,'' Int. J. Radiat. Oncol., Biol., Phys. **51**, 279–280 (2001).
[49] A. E. Nahum and D. M. Tait, ''Maximising local control by customising

dose prescription for pelvic tumours,'' in *Advanced Radiation Therapy—Tumor Response Monitoring and Treatment Planning*, edited by A. Breit (Springer-Verlag, Heidelberg, 1992), pp. 425–431.
[50] A. Terahara, A. Niemierko, M. Goitein, D. Finkelstein, E. Hug, N. Liebsch, D. O'Farrell, S. Lyons, and J. Munzenrider, ''Analysis of the relationship between tumor dose inhomogeneity and local control in patients with skull base chordoma,'' Int. J. Radiat. Oncol., Biol., Phys. **45**, 351–358 (1999).
[51] S. Levegrun, A. Jackson, M. J. Zelefsky, E. S. Venkatraman, M. W. Skwarchuk, W. Schlegel, Z. Fuks, S. A. Leibel, and C. C. Ling, ''Analysis of biopsy outcome after three-dimensional conformal radiation therapy of prostate cancer using dose-distribution variables and tumor control probability models,'' Int. J. Radiat. Oncol., Biol., Phys. **47**, 1245–1260 (2000).
[52] J. O. Deasy, ''Inter-patient and Intra-tumor radiosensitivity heterogeneity,'' in Ref. 46, pp. 363–381.
[53] R. A. Britten, A. J. Evans, M. J. Allalunis-Turner, A. J. Franko, and R. G. Pearcey, ''Intratumoral heterogeneity as a confounding factor in clonogenic assays for tumour radioresponsiveness,'' Radiother. Oncol. **39**, 145–153 (1996).
[54] B. Efron and R. J. Tibshirani, *An introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
[55] K. S. C. Chao, J. O. Deasy, J. Markman, J. Haynie, C. A. Perez, J. A. Purdy, and D. A. Low, ''A prospective study of salivary function sparing in patients with head and neck cancers receiving intensity-modulated or three-dimensional radiation therapy: initial results,'' Int. J. Radiat. Oncol., Biol., Phys. **49**, 907–916 (2001).
[56] J. O. Deasy, J. Markman, and K. S. C. Chao, ''Uncertainties in outcome predictions for treatment planning,'' Int. J. Radiat. Oncol., Biol., Phys. **49**, 907–916 (2001).
[57] J. Markman, J. O. Deasy, C. K. Chao, and D. Low, ''Bioeffect modeling of nonuniform dose distributions for paired organs,'' *Proceedings of the World Congress on Medical Physics and Bioengineering*, Chicago, IL, 2000 (CD-ROM).
[58] J. H. Zar, *Biostatistical Analysis* (Prentice–Hall, Englewood Cliffs, NJ, 1984).
[59] L. B. Marks *et al.*, ''The utility of SPECT lung perfusion scans in minimizing and assessing the physiologic consequences of thoracic irradiation,'' Int. J. Radiat. Oncol., Biol., Phys. **26**, 659–668 (1993).
[60] J. F. Fowler, ''Is repair of DNA strand break damage from ionizing radiation second-order rather than first-order? A simple explanation of apparently multiexponential repair,'' Radiat. Res. **152**, 124–136 (1999).
[61] P. A. Canney and W. T. Millar, ''Biphasic cellular repair and implications for multiple field radiotherapy treatments,'' Br. J. Radiol. **70**, 817–822 (1997).
[62] J. O. Deasy, J. F. Fowler, J. L. Roti Roti, and D. A. Low, ''Dose-rate effects in intensity modulated radiation therapy,'' Int. J. Radiat. Oncol., Biol., Phys. **51**, 400–401 (2001).
[63] J. Bijhold, ''Three-dimensional verification of patient placement during radiotherapy using portal images,'' Med. Phys. **20**, 347–356 (1993).
[64] D. Yan, D. Jaffray, and J. Won, ''A model to accumulate the fractionated dose in a deforming organ,'' Int. J. Radiat. Oncol., Biol., Phys. **44**, 665–675 (1999).
[65] W. Lu, ''A generalization of adaptive radiotherapy and the registration of deformable dose distribution,'' *XIIIth International Conference on The Use of Computers in Radiotherapy*, Heidelberg, Germany, 2000.
[66] J. Williamson, ''A deformable template approach to registration of serial CT studies for dose planning in intracavitary brachytherapy,'' in Ref. 65.
[67] J. Bijhold, J. V. Lebesque, A. M. Augustinus Hart, and R. E. Vijlbrirf, ''Maximizing setup accuracy using portal images as applied to a conformal boost technique for prostatic cancer,'' Radiother. Oncol. **24**, 261–271 (1992).
[68] D. Yan, D. Lockman, D. Brabbins, L. Tyburski, and A. Martinez, ''An off-line strategy for constructing a patient-specific planning target volume for image guided adaptive radiotherapy of prostate cancer,'' Int. J. Radiat. Oncol., Biol., Phys. **48**, 289–302 (2000).
[69] D. Yan and D. Lockman, ''Organ/patient positioning variation in external beam radiotherapy and its effects,'' Med. Phys. **28**, 593–602 (2001).
[70] J. Leong, ''Implementation of random positioning error in computerized radiation treatment planning systems as a result of fractionation,'' Phys. Med. Biol. **32**, 327–334 (1987).
[71] J. F. Fowler, ''The linear-quadratic formula and progress in fractionated radiotherapy,'' Br. J. Radiol. **62**, 679–694 (1989).
[72] G. W. Barensden, ''Dose fractionation, dose rate and iso-effect relationships for normal tissue response,'' Int. J. Radiat. Oncol., Biol., Phys. **8**, 1981–1997 (1982).
[73] B. Emami, J. Lyman, A. Brown, L. Coia, M. Goitein, J. E. Munzenrider, B. Shank, L. J. Solin, and M. Wesson, ''Tolerance of normal tissue to theraputic irradiation,'' Int. J. Radiat. Oncol., Biol., Phys. **21**, 109–122 (1991).
[74] H. Mameghan, R. Fisher, J. Mameghan, W. H. Watt, and A. Tynan, ''Bowel complications after radiotherapy for carcinoma of the prostate: The volume effect,'' Int. J. Radiat. Oncol., Biol., Phys. **18**, 315–320 (1990).
[75] D. P. Dearnaley, V. S. Khoo, A. R. Norman, L. Meyer, A. Nahum, D. Tait, J. Yarnold, and A. Howwich, ''Comparison of radiation side-effects of conformal and conventional radiotherapy in prostate cancer: a randomised clinical trial,'' Lancet **353**, 267–272 (1999).
[76] V. A. Benk, J. A. Adams, W. U. Shipley, M. M. Urie, P. L. McManus, J. T. Efrid, C. G. Willet, and M. Goitein, ''Late rectal bleeding following combined x-ray and proton high dose irradiation for patients with stages T3–T4 prostate carcinoma,'' Int. J. Radiat. Oncol., Biol., Phys. **26**, 551–557 (1993).
[77] C. A. Lawton *et al.*, ''Long-term treatment sequelae following external beam irradiation for adenocarcinoma of the prostate: analysis of RTOG studies 7506 and 7706,'' Int. J. Radiat. Oncol., Biol., Phys. **21**, 935–939 (1991).
[78] A. C. Hartford, A. Niemierko, J. A. Adams, M. M. Urie, and W. U. Shipley, ''Conformal irradiation of the prostate: estimating long term rectal bleeding using dose–volume histograms,'' Int. J. Radiat. Oncol., Biol., Phys. **36**, 721–730 (1996).
[79] T. E. Schultheiss, W. R. Lee, M. A. Hunt, A. L. Hanlon, R. S. Peter, and G. E. Hanks, ''Late GI and GU complications in the treatment of prostate cancer,'' Int. J. Radiat. Oncol., Biol., Phys. **37**, 3–11 (1997).
[80] L. J. Boersma, M. van den Brink, A. M. Bruce, T. Shouman, L. Gras, A. te Velde, and J. V. Lebesque, ''Estimation of the incidence of late bladder and rectum complications after high dose (70–78 Gy) conformal radiotherapy for prostate cancer, using dose–volume histograms,'' Int. J. Radiat. Oncol., Biol., Phys. **41**, 83–92 (1998).
[81] M. W. Skwarchuk, A. Jackson, M. J. Zelefsky, E. S. Venkatraman, D. M. Cowen, S. Levegrun, C. M. Burman, Z. Fuks, S. A. Leibel, and C. C. Ling, ''Late rectal toxicity after conformal radiotherapy of prostate cancer (I): Multivariate analysis and dose-response,'' Int. J. Radiat. Oncol., Biol., Phys. **47**, 103–113 (2000).
[82] A. Jackson, M. W. Skwarchuk, M. J. Zelefsky, D. M. Cowen, E. S. Venkatraman, S. Levegrun, C. M. Burman, G. J. Kutcher, Z. Fuks, S. A. Leibel, and C. C. Ling, ''Late rectal bleeding after conformal radiotherapy of prostate cancer (II): volume effects and dose volume histograms,'' Int. J. Radiat. Oncol., Biol., Phys. **49**, 685–698 (2001).
[83] A. Jackson, M. W. Skwarchuk, and S. Levegrun, ''Volume effects in external beam treatments of prostate cancer'' (extended abstract), in (*Proceedings of the 5th International Symposium on 3D Conformal Radiation Therapy and Brachytherapy*, edited by H. I. Amols and H. J. Zelefsky, Memorial Sloan–Kettering Cancer Center, New York, NY, 1–3 June 2000.
[84] J. D. Fenwick, V. S. Khoo, A. E. Nahum, B. Sanchez-Nieto, and D. P. Dearnaley, ''Correlations between dose–surface histograms and the incidence of long-term rectal bleeding following conformal or conventional radiotherapy treatment of prostate cancer,'' Int. J. Radiat. Oncol., Biol., Phys. **49**, 473–480 (2001).
[85] L. J. Boersma, E. M. Damen R. W. de Boer, S. H. Muller, R. A. Valdes Olmos, C. A. Hoefnagel, C. M. Roos, N. van Zandwijk, and J. V. Lebesque, ''A new method to determine dose-effect relations for local lung-function changes using correlated SPECT and CT data,'' Radiother. Oncol. **29**, 110–116 (1993).
[86] L. B. Marks, M. T. Munley, D. P. Spencer, G. W. Sherouse, G. C. Bentel, J. Hoppenworth, M. Chew, R. J. Jaszczak, R. E. Coleman, and L. R. Prosnitz, ''Quantification of radiation-induced regional lung injury with perfusion imaging,'' Int. J. Radiat. Oncol., Biol., Phys. **38**, 399–409 (1997).
[87] L. B. Marks, D. Hollis, M. Munley, G. Bentel M. Garipagaoglu, M. Fan, J. Poulson, R. Clough, G. Sibley, R. E. Coleman, and R. Jaszczak, ''The role of lung perfusion imaging in predicting the direction of radiation-induced changes in pulmonary function tests,'' Can. Mineral. **88**, 2135–2141 (2000).
[88] Y. Seppenwoolde, S. H. Muller, J. C. Theuws, P. Baas, J. S. Belderbos, L.

J. Boersma, and J. V. Lebesque, ''Radiation dose-effect relations and local recovery in perfusion for patients with non-small-cell lung cancer,'' Int. J. Radiat. Oncol., Biol., Phys. **47**, 681–690 (2000).
[89] J. M. Balter, R. K. Ten Haken, T. S. Lawrence, K. L. Lam, and J. R. Robertson, ''Uncertainties in CT-based treatment plans due to patient breathing,'' Int. J. Radiat. Oncol., Biol., Phys. **36**, 167–174 (1996).
[90] J. G. Armstrong, M. J. Zelefsky, S. A. Leibel, C. Burman, C. Han, L. B. Harrison, G. J. Kutcher, and Z. Y. Fuks, ''Strategy for dose escalation using 3-dimensional conformal radiation therapy for lung cancer,'' Ann. Oncol. **6**, 693–697 (1995).
[91] L. A. Dawson, C. McGinn, D. Normolle, R. K. Ten Haken, W. Ensminger, and T. S. Lawrence, ''Escalated focal liver radiation and hepatic artery fluorodeoxyuridine for unresectable liver malignancies,'' J. Clin. Oncol. **18**, 2210–2218 (2000).
[92] L. A. Dawson, R. K. Ten Haken, and T. S. Lawrence, ''Partial irradiation of the liver,'' Semin Radiat. Oncol. **11**, 234–239 (2001).
[93] M. V. Graham, J. A. Purdy, B. Emami, J. W. Matthews, and W. B. Harms, ''Preliminary results of a prospective trial using three dimensional radiotherapy for lung cancer,'' Int. J. Radiat. Oncol., Biol., Phys. **33**, 993–1000 (1995).
[94] J. A. Hayman, M. K. Martel, R. K. Ten Haken, D. P. Normolle, R. F. Todd III, J. F. Littles, M. A. Sullivan, P. W. Possert, A. T. Turrisi, and A. S. Lichter, ''Dose escalation in non-small cell lung cancer using 3-dimensional conformal radiation therapy: update of a phase I trial,'' J. Clin. Oncol. **19**, 127–136 (2001).
[95] B. L. Kwa, J. V. Lebesque, J. C. Theuws, L. B. Marks, M. T. Munley, G. Bentel, D. Oetzel, U. Spahn, M. V. Graham, R. E. Drzymala, J. A. Purdy, A. S. Lichter, M. K. Martel, and R. K. Ten Haken, ''Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients,'' Int. J. Radiat. Oncol., Biol., Phys. **42**, 1–9 (1998).
[96] L. B. Marks, M. Fan, R. Clough, M. Munley, G. Bentel, R. E. Coleman, R. Jaszczak, D. Hollis, and M. Anscher, ''Radiation-induced pulmonary injury: symptomatic versus subclinical endpoints,'' Int. J. Radiat. Oncol., Biol., Phys. **76**, 469–475 (2000).
[97] M. K. Martel, Ten Haken, M. B. Hazuka, A. T. Turrisi, B. A. Fraass, and A. S. Lichter, ''Dose–volume histogram and 3-D treatment planning evaluation of patients with pneumonitis,'' Int. J. Radiat. Oncol., Biol., Phys. **28**, 575–581 (1994).
[98] C. J. McGinn, R. K. Ten Haken, W. D. Ensminger, S. Walker, S. Wang, and T. S. Lawrence, ''The treatment of intrahepatic cancers with radiation doses based on a normal tissue complication probability model,'' J. Clin. Oncol. **16**, 2246–2252 (1998).
[99] D. Oetzel, P. Schraube, F. Hensley, G. Sroka-Perez, M. Menke, and M. Flentje, ''Estimation of pneumonitis risk in three-dimensional treatment planning using dose-volume histogram analysis,'' Int. J. Radiat. Oncol., Biol., Phys. **33**, 455–460 (1995).
[100] J. M. Robertson, R. K. Ten Haken, A. T. Turrisi, M. B. Hazuka, M. K. Martel, J. F. Littles, F. J. Martinez, I. Francis, L. Quint, and A. S. Lichter, ''Dose escalation for non-small cell lung cancer using conformal radiation therapy,'' Int. J. Radiat. Oncol., Biol., Phys. **37**, 1079–1085 (1997).
[101] A. Jackson, C. C. Ling, C. M. Burman, E. S. Venkatraman, J. Bass, A. Raben, S. A. Leibel, and Z. Fuks, ''A conformal dose escalation protocol for the treatment of tumors in the lung or other normal tissues with parallel architecture,'' Med. Phys. **23**, 1495 (1996).
[102] R. K. Ten Haken, M. K. Martel, M. L. Kessler, M. B. Hazuka, T. S. Lawrence, J. M. Robertson, A. T. Turrisi, and A. S. Lichter, ''Use of Veff and iso-NTCP in the implementation of dose escalation protocols,'' Int. J. Radiat. Oncol., Biol., Phys. **27**, 689–695 (1993).
[103] A. Jackson, R. K. Ten Haken, J. M. Robertson, M. L. Kessler, G. J. Kutcher, and T. S. Lawrence, ''Analysis of clinical complication data for radiation hepatitis using a parallel architecture model,'' Int. J. Radiat. Oncol., Biol., Phys. **31**, 883–891 (1995).
[104] S. A. Roberts and J. H. Hendry, ''The delay before onset of accelerated tumour cell repopulation during radiotherapy: a direct maximum-likelihood analysis of a collection of worldwide tumour-control data,'' Radiother. Oncol. **29**, 69–74 (1993).
[105] R. K. Ten Haken, T. S. Lawrence, D. L. McShan, R. J. Tesser, B. A. Fraass, and A. S. Lichter, ''Technical considerations in the use of 3-D beams in the abdomen,'' Radiother. Oncol. **22**, 19–28 (1991).
[106] T. S. Lawrence, R. J. Tesser, and R. K. Ten Haken, ''An application of dose volume histograms to treatment of intrahepatic malignancies with radiation,'' Int. J. Radiat. Oncol., Biol., Phys. **19**, 1041–1047 (1990).
[107] T. S. Lawrence, L. M. Dworzanin, S. C. Walker-Andrews, J. C. Andrews, R. K. Ten Haken, I. S. Wollner, A. S. Lichter, and W. D. Ensminger, ''Treatment of cancers involving the liver and porta hepatis with external beam irradiation and intraarterial hepatic fluorodeoxyuridine,'' Int. J. Radiat. Oncol., Biol., Phys. **20**, 555–561 (1991).
[108] J. T. Lyman, ''Complication probability as assessed from dose volume histograms,'' Radiat. Res. **104**, S13–S19 (1985).
[109] T. S. Lawrence, R. K. Ten Haken, M. L. Kessler, J. M. Robertson, J. T. Lyman, M. L. Lavigne, M. B. Brown, D. J. DuRoss, J. C. Andrews, W. D. Ensminger, and A. S. Lichter, ''The use of 3-D dose volume analysis to predict radiation hepatitis,'' Int. J. Radiat. Oncol., Biol., Phys. **23**, 781–788 (1992).
[110] G. J. Kutcher and C. Burman, ''Calculation of complication probability factors for nonuniform normal tissue irradiation: the effective volume method,'' Int. J. Radiat. Oncol., Biol., Phys. **16**, 1623–1630 (1989).
[111] J. Hanley, M. M. Debois, D. Mah, G. S. Mageras, A. Raben, K. Rosenzweig, B. Mychalczak, L. H. Schwartz, P. J. Gloeggler, W. Lutz, C. C. Ling, S. A. Leibel, Z. Fuks, and G. J. Kutcher, ''Deep inspiration breath-hold technique for lung tumors: the potential value of target immobilization and reduced lung density in dose escalation,'' Int. J. Radiat. Oncol., Biol., Phys. **45**, 603–611 (1999).
[112] J. W. Wong, M. B. Sharpe, D. A. Jaffray, V. R. Kini, J. M. Robertson, J. S. Stromberg, and A. A. Martinez, ''The use of active breathing control (ABC) to reduce margin for breathing motion,'' Int. J. Radiat. Oncol., Biol., Phys. **44**, 911–919 (1999).
[113] J. M. Balter, K. L. Lam, C. J. McGinn, T. S. Lawrence, and R. K. Ten Haken, ''Improvement of CT-based treatment planning models of abdominal targets using static exhale imaging,'' Int. J. Radiat. Oncol., Biol., Phys. **41**, 939–943 (1998).
[114] R. K. Ten Haken, J. M. Balter, L. H. Marsh, J. M. Robertson, and T. L. Lawrence, ''Potential benefits of eliminating planning target volume expansions for patient breathing in the treatment of liver tumors,'' Int. J. Radiat. Oncol., Biol., Phys. **38**, 613–617 (1997).
[115] A. E. Lujan, E. W. Larsen, J. M. Balter, and R. K. Ten Haken, ''A method for incorporating organ motion due to breathing into 3D dose calculations,'' Med. Phys. **26**, 715–720 (1999).
[116] A. E. Lujan, R. K. Ten Haken, E. W. Larsen, and J. M. Balter, ''Quantization of setup uncertainties in 3D dose calculations,'' Med. Phys. **26**, 2397–2402 (1999).
[117] N. Choi, J. E. Herndon II, J. Rosenmann, R. W. Carey, C. T. Chung, S. Bernard, L. Leone, S. Seagren, and M. Green, ''Phase I study to determine the maximum-tolerated dose of radiation in standard daily and hyperfractioned-accelerated twice-daily radiation schedules with concurrent chemotherapy for limited stage small-cell lung cancer,'' J. Clin. Oncol. **16**, 3528–3537 (1998).
[118] C. R. Thomas, Jr., D. J. Giroux, K. J. Stelzer, J. B. Craig, L. R. Laufman, S. A. Taylor, J. W. Goodwin, J. J. Crowley, and R. B. Livingston, ''Concurrent cisplatin prolonged oral etoposide, and vincristine plus chest and brain irradiation for limited small cell lung cancer: A phase II study of the southwest oncology group (SWOG-9229),'' Int. J. Radiat. Oncol., Biol., Phys. **40**, 1045–1047 (1998).
[119] M. V. Graham, ''Predicting radiation response,'' Int. J. Radiat. Oncol., Biol., Phys. **39**, 561–562 (1997).
[120] L. Marks, M. T. Munley, G. C. Bentel, S.-M. Zhou, D. Hollis, C. Scarfone, G. S. Sibley, F.-M. Kong, R. Jirtle, R. Jaszczak, R. E. Coleman, V. Tapson, and M. Anscher, ''Physical and biological predictors of changes in whole lung function following thoracic irradiation,'' Int. J. Radiat. Oncol., Biol., Phys. **39**, 563–570 (1997).
[121] M. V. Graham, J. A. Purdy, B. Emami, W. Harms, W. Bosch, M. A. Lockett, and C. A. Perez, ''Clinical dose–volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC),'' Int. J. Radiat. Oncol., Biol., Phys. **45**, 323–329 (1999).
[122] M. Langer and P. Kijewski, ''CCRT for non-small cell lung cancer: sensitivity of clinical gains to organ tolerance restrictions,'' Int. J. Radiat. Oncol., Biol., Phys. **22**, 325–332 (1992).
[123] C. Ha, P. J. Kijewski, and M. P. Langer, ''Gain in target dose from using computer controlled radiotherapy (CCRT) in the treatment of non-small cell lung cancer.'' Int. J. Radiat. Oncol., Biol., Phys. **26**, 335–356 (1991).
[124] M. V. Graham, J. A. Purdy, B. Emami, W. Harms, W. Bosch, M. A. Lockett, and C. A. Perez, ''Clinical dose–volume histogram analysis for pneumonitis after 3D treatment for non-small lung cancer,'' Int. J. Radiat. Oncol., Biol., Phys. **45**, 323–329 (1999).

[125] M. Langer, P. Kijewski, R. Brown, and C. Ha, ''The effect on minimum tumor dose of restricting target-Dose inhomogeneity in optimized three-dimensional treatment of lung cancer,'' Radiother. Oncol. **21**, 245–256 (1991).
[126] M. V. Graham, J. W. Mathews, W. B. Harms, B. Emami, H. S. Glazer, and J. A. Purdy, ''Three-dimensional radiation treatment planning study for patients with carcinoma of the lung,'' Int. J. Radiat. Oncol., Biol., Phys. **29**, 1105–1117 (1994).
[127] H. H. Liu, I. I. Rosen, N. A. Janjan, and A. Pollack, ''Treatment planning optimization based on response-surface modeling of cost function versus multiple constraints,'' *CD-ROM Proceedings of the World Congress on Medical Physics and Biomedical Engineering*, 23–28 July 2000.
[128] M. Langer, V. Thai, and L. Papiez, ''Sensitivity analysis of tolerance limits for dose-escalation,'' in Ref. 127.
[129] M. Langer, S. Morrill, and R. Lane, ''A test of the claim that plan rankings are determined by relative complication and tumor control probabilities,'' Int. J. Radiat. Oncol., Biol., Phys. **41**, 451–457 (1998).
[130] G. B. Dantzig, in ''Convex programming,'' *Linear Programming and Extensions* (Princeton University Press, Princeton, NJ), Chap. 24, pp. 471–498.
[131] E. Fermi, ''The nucleus'' Phys. Today **5**, 6–9 (1952).