Applying Probabilistic Decision Models to Clinical Trial Design

Wade P Smith, Ph.D.[1], **Mark H Phillips, Ph.D.**[1,2]
[1]**Department of Radiation Oncology, University of Washington, Seattle, WA;** [2]**Biomedical Informatics and Medical Education, University of Washington, Seattle, WA**

Abstract

Clinical trial design most often focuses on a single or several related outcomes with corresponding calculations of statistical power. We consider a clinical trial to be a decision problem, often with competing outcomes. Using a current controversy in the treatment of HPV-positive head and neck cancer, we apply several different probabilistic methods to help define the range of outcomes given different possible trial designs. Our model incorporates the uncertainties in the disease process and treatment response and the inhomogeneities in the patient population. Instead of expected utility, we have used a Markov model to calculate quality adjusted life expectancy as a maximization objective. Monte Carlo simulations over realistic ranges of parameters are used to explore different trial scenarios given the possible ranges of parameters. This modeling approach can be used to better inform the initial trial design so that it will more likely achieve clinical relevance.

Introduction

The design of a clinical trial can have a large impact on the progress of medical practice. A number of competing factors need to be considered, such as cohort size, ability to recruit subjects, overall cost, expected size of effect and expected impact of the outcome of the trial. While these and other factors are considered in the design of any trial, there is continuing controversy over how best to do it. One persistent question is whether a clinical trial is solely for the purpose of inference or whether it is a decision problem (for a concise synopsis see Spiegelhalter *et al*[1]).

Traditionally, clinical trial designs are a combination of relatively straight-forward statistics and clinical judgment. The latter informs the design by providing an estimate of the effect size that will alter clinical practice, i.e. there is no point in conducting a trial whose positive outcome will not change how physicians practice. The factors to be considered typically include things like possible increase in side-effects, ease of administering the treatment and cost. If it seems reasonable that the effect to be measured will achieve such a magnitude, then the number of trial subjects is calculated using classical statistics to minimize false positive and false negative trial results. In this approach to trial design, one seeks to *infer* the true value of the treatment difference between the two arms from the sample obtained through the clinical trial.

When a trial is viewed as an inference problem, the focus is on determining whether the statistics of the trial provide enough certainty to achieve a significant p-value with one, or possibly several related, outcomes. For example, in cancer therapy the common outcomes studied are overall survival, disease-free survival, and local recurrence[2]. Classic trial design then continues by determining the likelihood function for the chosen outcome, applying the parameters of the control group, agreeing on a clinically relevant difference between new and old treatments, and finally, calculating numbers of subjects given acceptable thresholds of Type I and II errors. Recently, Bayesian inference in clinical trial design has become more common and it can be used for two different purposes. First, using Bayes' theorem allows for the inspection of the most probable outcomes after only some fraction of the expected number of patients are accrued. This can help to direct resources more efficiently later in the trial and/or to insure that any potential harm is minimized. The other use of Bayesian statistics is to quantify the magnitude of the treatment effect rather than just testing the null hypothesis.

As mentioned, trial design needs to take into account aspects of the new procedure or treatment beyond the main outcome of interest. For example, a clinical trial of a new cancer therapy will be designed to measure the difference in survival at a given time point. However, the future use of the new therapy will also depend on whether more or different complications result, how difficult it is for the patient and/or provider to administer the treatment, and perhaps how much the new treatment costs. These conflicting considerations characterize problems that are best approached using decision theory.

If trial design is viewed as a formal decision problem, then a loss (or *utility* or *cost*) function is needed to quantify

the effects, and the expectation value of the utility (or cost) is used in the decision making process[3]. This function is designed to capture the many different aspects of the outcome–the response of the disease, the induced health state, and the feasibility and cost of the new procedure or treatment. The fact that the uncertainties inherent in the problem result in the use of of expectation values of the utility indicates that a model of the probabilities of the outcomes is needed. While this adds somewhat to the complexity of framing and quantifying the medical environment compared to the inference problem, it has the benefit of being able to incorporate multiple, competing outcomes. Thus, one obtains a much more complete and clinically realistic view with the aim of designing a trial that will more likely have an impact in clinical decision making.

In this paper, we take the more comprehensive view and model clinical trial design as a decision problem. We combine several probabilistic modeling approaches in order to capture the uncertainties in the disease and treatment response processes and the inhomogeneities within the subject cohorts. The separation of these two sources of variability is an explicit part of our model which is needed to account for their potential effects. In addition, it allows us to generalize our model that was designed for clinical trial design and apply it to clinical decision making, as well. Our model also makes use of a more general utility function, namely quality adjusted life years (QALY). This function captures the interplay between cure and complication as well as integrating the outcomes over time.

Overall, our approach offers several advantages over current methods. It allows us to simulate different measured outcomes, thereby giving quantitative guidance as to whether a given trial design is likely to result in clinically useful information. In addition, by calculating the expected value of information regarding different variables, the trial can be designed to either control for certain variables or, prior to initiating the trial, more information can be obtained prior to the trial. Finally, it should be mentioned that our method focuses on the effect size. Traditional trials follow the "null hypothesis significance testing" approach which fails to give any information about the effect size. As our analysis shows, realistic modeling of effect size is a key element to a successful trial.

Methods

We have applied our method to clinical trials for HPV-positive (HPV+) oropharyngeal cancer. Subgroup analysis of large clinical trials show that HPV+ patients have significantly improved survival compared to HPV-negative (HPV-) patients under standard radiation therapy protocols[4]. A number of non-inferiority clinical trials have been initiated to determine if lower doses of radiation can be equally efficacious, the goal being to decrease the incidence of severe complications due to the radiation therapy[5]. Thus, the trial design must include a measure of the difference in survival at some relevant time and the differences in complications. In this case, the new treatment uses fewer treatments which is acknowledged to be a benefit to the patient and to incur less cost, so there is no need to model any tradeoff in that area.

Model framework

The structure of the model was obtained through in depth discussions with experts in the field of head and neck cancer radiation therapy and from previous publications[6]. We constructed an influence diagram[7] (Figure 1) that incorporates the relevant variables with respect to tumor control and complications to nearby critical structures, along with the decision variables (radiation dose and technique). An influence diagram is a Bayesian network (BN) that includes decision nodes and a utility node. In Figure 1, rectangles denote decision (also known as *action*) nodes; ovals represent chance nodes; and diamonds represent utility nodes. Color coding of the chance nodes indicate the details of the implementation of joint probability tables. Green nodes indicate prior distributions; red nodes indicate model-generated probabilities; and yellow indicate normal chance nodes.

In a conventional influence diagram, the expected utility (denoted by a diamond-shaped node) is calculated by means of the relevant probabilities obtained from the BN and a utility function. We have expanded the notion of expected utility from the standard form[8]. Instead of using a single function to map preferences and probabilities to an expected utility, we use a Markov model to calculate quality adjusted life years (QALY) (Figure 2)[9]. The BN provides the distribution of initial states to the Markov model. The Markov model contains states for status-post RT, recurrent disease, Gr2+ xerostomia, Gr2+ dysphagia, both toxicities, and death. To each state is assigned a utility, and QALYs are accumulated as the Markov model is evaluated annually using a cohort simulation[10,11].

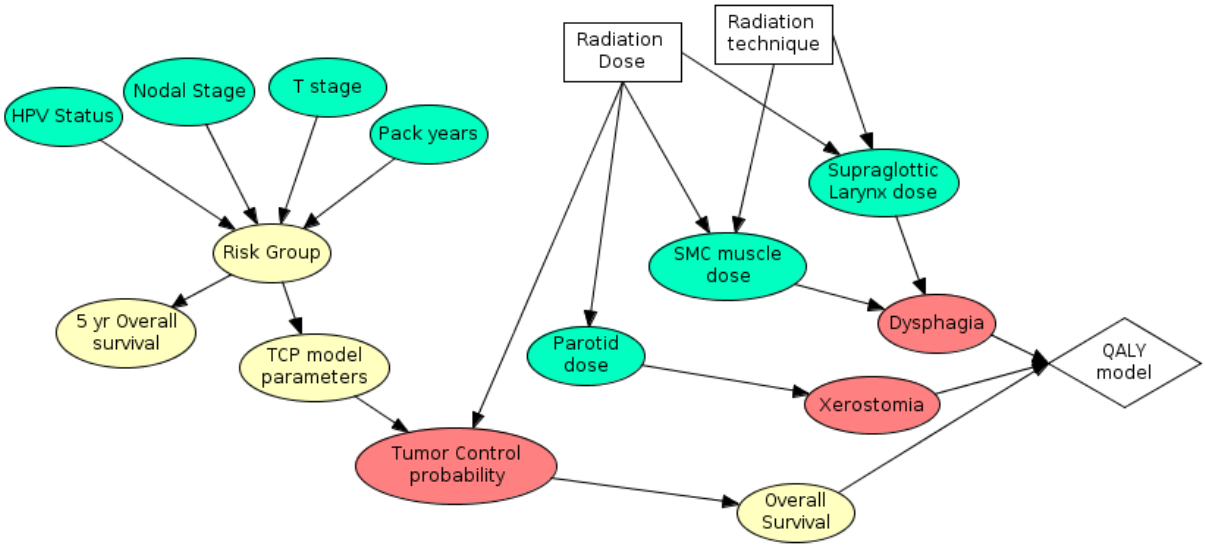

Figure 1: Influence diagram modeling the treatment and response of HPV+ oropharyngeal cancer.

In a conventional influence diagram, the optimal action can be solved in a direct fashion[7]. However, the use of the the Markov model complicates this process; therefore, we have used a Monte Carlo process to provide values of the probabilities from the BN and to evaluate the Markov model and calculate QALY. These values are then used to choose the optimal action. This approach also allows us to perform a sensitivity analysis as part of the QALY calculation in a natural way. For any given distribution of a variable, the Monte Carlo process samples the distribution that describes the variabilities inherent in its value.

Tumor Control Probability

Modeling the tumor control probability (TCP) is one the most crucial steps. The non-inferiority trials being planned are based on the results reported by Ang *et al*[4] for HPV+ patients, and our model relies on the same. They stratified patients based on various risk factors, such as smoking and clinical pathology variables (see Figure 1). Most HPV+ patients were in the low risk category (smoking history being a deciding factor), and our current model focuses solely on that group. Ang *et al* reported a 3 year survival rate of 82.4% [95% CI, 77.2 to 87.6][4]. Given this variance, a clinically meaningful difference in survival would need to be 5% or greater. The dose reduction trials are looking at the difference in survival between a tumor dose of 70 Gy and a dose of 55 Gy. We modeled three different scenarios, namely that a dose reduction from 70 Gy to 55 Gy would yield either a 7%, 10% or 15% reduction in survival.

TCP is nearly always modeled as a logistic curve with two free parameters, the dose at which TCP = 0.5 [D_{50}] and the slope at that point, γ_{50}. We modeled these two parameters as beta functions from which the Monte Carlo process sampled. For each of the three scenarios, the TCP was calculated using the sampled values and only those values were used for which TCP(70 Gy) = 82% and the difference in survival at TCP(55 Gy) $\in [7, 10, 15]$%. The means and probabilistic sensitivity analysis ranges (95%) were 34.7, 22.5 –45.7 Gy and 0.55, 0.24 – 0.98 for D_{50} and γ_{50}, respectively.

Normal Tissue Complication Probability: xerostomia

Normal tissue complication probabilities (NTCP) are calculated by means of logistic functions that are fitted to the data for given grades (or range of grades) of complications to a given organ. Xerostomia ("dry mouth") occurs when the parotid glands suffer radiation injury. Such complications are common in oropharyngeal cancer with 40-60% of patients suffering significant sequelae. In contrast to the tumor in which the planned dose is uniformly distributed throughout the volume, normal tissues receive a distribution of doses. The actual distribution is most strongly influ-

enced by a particular patient's anatomy, that is, the geometric relationship between the tumor volume and the organ at risk. NTCP models usually use some summary value, e.g. mean dose, although more complex, radiobiologically-based models are also used[12].

To calculate the probability of xerostomia, we used a model published by Miah *et al*[13], which uses the mean dose to the parotid gland. This model used non-linear logistic regression analysis to determine the probability of xerostomia from a model with two parameters, D50, the mean dose which induces grade ≥ 2 xerostomia in 50% of patients, and k, the additional rate of xerostomia for each additional Gy at D50. The mean and range on these parameters for the Monte Carlo sensitivity analysis is 42.9 Gy [29-56 Gy] and 1.7 [0.5-2.8]. We also used the Monte Carlo process to sample a distribution of mean parotid doses that were obtained from the treatment planning database of the Department of Radiation Oncology, University of Washington. The mean and range of the parotid mean dose was 21.4 [6–38.9] Gy.

Normal Tissue Complication Probability: dysphagia

In our influence diagram probabilities for dysphagia come from an NTCP model developed from a prospective study[14]. Their work finds values for the logistic function regression coefficients of 0.04 ± 0.02 and 0.06 ± 0.02 for the SGL and SPCM, respectively; these are used as distributions in the Monte Carlo sensitivity analysis. Their model predicts swallowing dysfunction as a function of mean doses to the supraglottic larynx (SGL) and superior pharyngeal constrictor muscle (SPCM). Possible ranges for mean doses to these structures are dependent on the radiation planning technique employed which varies across clinical practices. We used a range of mean doses appropriate for an IMRT planning technique which attempts to spare the SGL and SPCM[15], as this technique is likely to be employed by physicians considering prescription dose de-escalation in order to reduce toxicity. The means and ranges were 52.5 Gy [29.6 – 67.4 Gy] and 63.0 Gy [34.5 – 69.8 Gy] for the SGL mean dose and SPCM mean dose, respectively.

Utility values

Multiattribute utilities for complex health states are not easy to obtain and can depend on the method of elicitation. Meregaglia *et al* reviewed a wide range of publications in the area of utilities for health states relevant to head and neck cancers[16]. We used the values collected by Ramaekers *et al* for our model[17]. In our model, we sampled from distributions that we feel mainly represent heterogeneity between patients, though an interpretation that the distributions actually represent uncertainty regarding methods is not unassailable.

Monte Carlo simulations

Probabilistic sensitivity analyses were conducted using a Monte Carlo sampling technique with 50,000 patients[10]. By this we mean that a single patient was simulated by selecting at random from each of the distributions in the diagram. The probability of selection of any given value is determined by the functional form of distributions that we used. For example, a normally distributed variable with a given mean and standard deviation would be sampled such that by the end of the analysis, the distribution of values used would match the initial normal distribution. Since each variable is sampled independently, one ends up with 50,000 "patients" whose characteristics are distributed as expected in a heterogeneous population. For each run of 50,000 samples, the QALYs for each patient were calculated and histograms of the results plotted. Separate runs were calculated for each of the different scenarios that represent possible trial outcomes (current knowledge, survival reductions of 7%, 10% and 15%) and for each of the four actions (delivery of 70, 65, 60 and 55 Gy).

Results

The results of our simulation using this model are in the form of: (a) a table of expected value of perfect and partial perfect information, (b) histograms of the QALY's achieved under the different possible trial outcomes and for different delivered tumor doses, and (c) a plot of the fraction of patients who would benefit from each of the possible decisions (tumor doses) as a function of true survival differences.

Table 1 contains the values of expected value of perfect information (EVPI) and expected value of partial perfect

information (EVPPI)[18]. The column EVPI presents the results when all variables are assumed known, and places an upper bound on what can be achieved by acquiring perfect information on only a subset of variables (EVPPI). The five columns under the heading EVPPI present the results of partial perfect information, i.e. each variable listed represents the QALYs gained if the given variable is known exactly.

Value of information [QALYs per 10,000]

	EVPI	EVPPI				
scenarios		utilities	TCP	NTCP(xerostomia)	NTCP(dysphagia)	Organ-at-risk doses
current knowledge	617	112	66	0	14	0
15%	714	239	70	0	94	0
10%	1175	564	172	24	437	5
7%	1239	653	31	37	614	22

Table 1: Value of information calculations for expected value of perfect information (EVPI) and partial perfect information (EVPPI) (columns 3–7) for different variables. VOI is calculated for four different scenarios: only using our current knowledge, a reduction in cure rate of 15% under a dose reduction of 77 to 55 Gy, a reduction in cure rate of 10% and a reduction in cure rate of 7%.

Figure 2 shows the distributions of QALYs for a range of doses between 70 and 55 Gy. We included 65 and 60 Gy in order to illustrate that this method could be used to guide selection of which doses are chosen as experimental arms. Here the possible trial result that dose de-escalation to 55 Gy results in a less than 7% reduction in overall survival has been entered into the model in order to evaluate which uncertainties contribute the most to the decision uncertainty for a particular patient.

Note that all four panes of Figure 2 are calculated on the same modeled cohort, but with different prescription doses. For some individuals, the expected benefit increases, and for some it decreases, depending on the individual's probabilities of developing side effects, their preference for the various health states, and their expected probability of survival. The standard deviations of the distributions are reduced as the prescription dose is lowered. In order to illustrate why this occurs we performed a subset analysis on the patients whose QALYs for 70 Gy are greater than the maximum (16.7 QALYs) expected benefit calculated for the 55 Gy arm. None of the individuals in this subset can benefit from a dose reduction to 55 Gy. In this group the average utilities for both side effects were higher than the cohort average, and the average doses to all critical structures were lower than the cohort average, thereby lowering the probability of developing side effects. Patients at these extreme values, are expected to derive the most benefit from maximizing life expectancy, which occurs at the highest dose. Similarly, the group of modeled patients whose expected QALYs for 70 Gy were less than the minimum expected for 55 Gy would benefit from a dose reduction. Patients in this modeled cohort assign lower utilities to health states with toxicity, and also had a higher probability than the cohort average for developing a side effect. This group benefits the most from avoiding toxicity, which occurs at the lowest dose.

Figure 3 shows the distributions of QALYs for the current state of knowledge and for range of possible trial results for the de-escalated dose to 55 Gy. As the difference in overall survival between 70 Gy and 55 Gy decreases, the average population QALYs increases due to the higher survival rate on the 55 Gy arm. The probability of toxicity events is the same for these 4 scenarios, since all are calculated for 55 Gy, and the toxicity and tumor control probabilities are independent probabilities.

The results of the Monte Carlo simulation allow for the calculation of the proportion of patients who would achieve the highest number of QALYs for a given strategy (tumor dose). Figure 4 plots the proportions of patients who benefit from a given dose as a function of the true difference in survival at doses of 70 and 55 Gy, i.e. TCP(70) - TCP(50). At large differences, the decrease in complications cannot overcome the relative importance improved survival. The ordering of the curves at high doses reflect the nature of the logistic TCP curves. At small differences, the advantage of complication sparing is more important as the survival values are closer in value.

For smaller reductions in overall survival (the left side of Fig. 4 or the final row of Table 1) the EVPI increases since decreased doses offer more QALYs to a greater proportion of the population, and perfect knowledge of other model

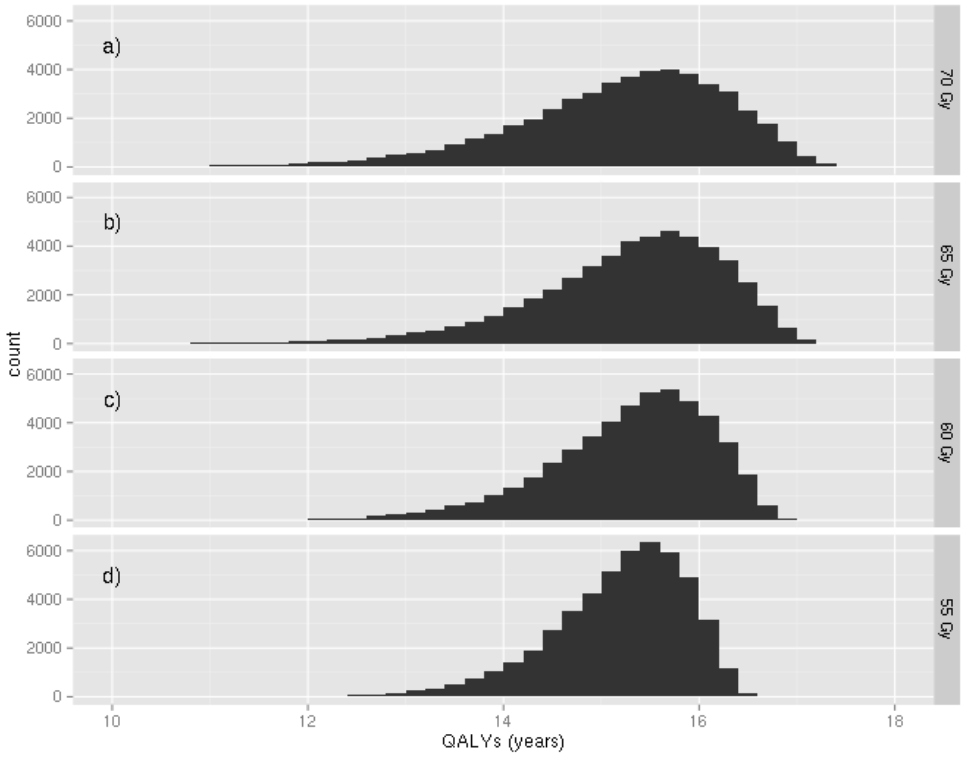

Figure 2: Histograms of QALY's calculated under the hypothesis that a reduction in tumor dose from 70 to 55 Gy would result in a 7% reduction in survival. Calculations were performed for four different values of the delivered tumor dose: 70, 65, 60, 55 Gy.

parameters would allow the optimal choice for each individual. With the current state of knowledge, acquiring additional information about any one set of parameters does not have as significant an impact on the decision made, since the remaining variability has a large effect. However, for trial results where dose reduction results in smaller reduction in survival, then acquiring additional information about patient preferences or dysphagia NTCP model parameters has greater potential to change the decision.

Discussion

We have presented a model that is based on two major premises. First, the design of clinical trials should be considered as a decision problem, not merely as an inference problem. Second, the utility function needs to include multiple attributes as well as integration over time. The former was addressed by means of a probabilistic knowledge representation in the form of an influence diagram. The latter was accounted for by using a Markov model to calculate quality adjusted life years. This decision rendered the normal approach to solving the influence diagram for the optimal decision difficult to achieve; therefore, we used a Monte Carlo sampling method.

We applied this clinical trial design model to the case of non-inferiority trials for HPV+ oropharyngeal cancer. Radiation therapy (with and without chemotherapy) of oropharyngeal cancer leads to significant sequelae in a large fraction of patients. The most recent large clinical trial identified patients who were HPV+ as having better survival probability than HPV- patients[4]. Given the complication profile, the question arises as whether a reduced radiation dose for HPV+ patients might yield similar survival benefits with reduced complications when compared to current dose protocols. With this in mind, a number of clinical dose de-escalation trials have begun[5].

Classic radiation biology theory holds that a reduction of dose on the order of over 20% will result in a decrease in cell killing, with a concommitant decrease in tumor-related outcomes. What is unknown is whether the dose-response

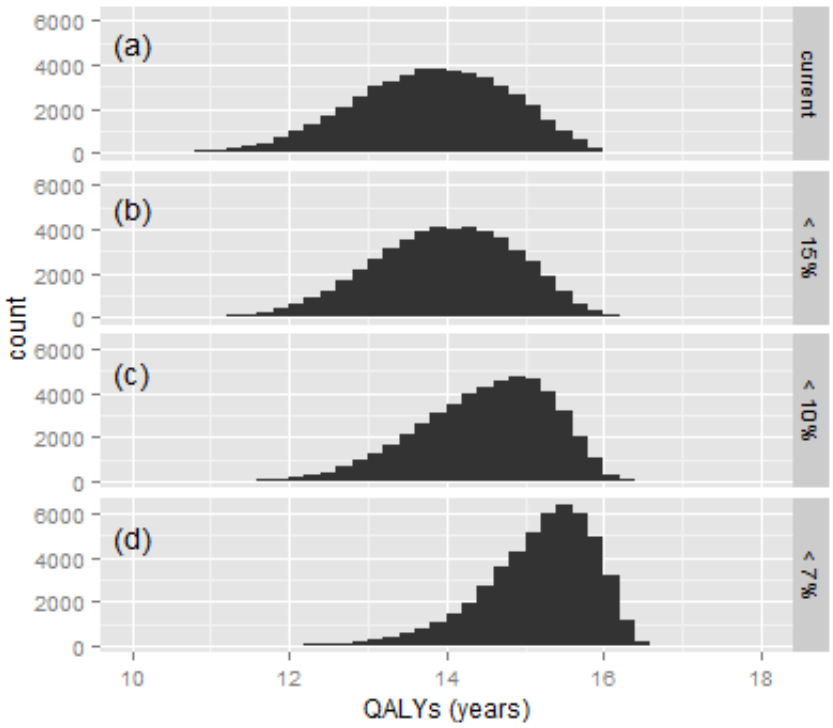

Figure 3: Histograms of QALY's achieved if the delivered tumor dose is 55 Gy under four different possible trial outcomes: (a) current protocol with no additional knowledge, (b) 55 Gy results in a 15% reduction in TCP, (c) 55 Gy results in a 10% reduction, and (d) 55 Gy results in a 7% reduction.

curve is so flat that the outcome is essentially the same. There are two ways that this could come about. Since our data are statistical in nature, non-inferiority can imply confirmation of the standard null hypothesis between the current and the lower dose protocols. The second way that equivalence could be established is to acknowledge that clinical decisions incorporate consideration of both cure and complication, i.e. viewing the trial as a decision problem.

Consider the first possibility, namely that a clinical trial need only confirm the null hypothesis. Ang *et al* report a standard deviation of 2.5%[4]. Since we are looking at the difference in survival, the ability to distinguish two different mean survival values depends on the sum of the variances. Therefore, in this case, a trial result that could reliably reject the null hypothesis at a level of 0.05 would imply a difference in means of approximately 7%. This consideration led us to our first scenario in which a trial based on inference is on the threshold of significance, that is, able to achieve a p-value of 0.05 or better. Our other scenarios (differences of 10% and 15%) were designed to explore the possibility that reduction in complications could compensate for a larger decrease in survival.

The histograms (Figures 2 and 3) show the changes in the distributions of QALY's obtained for different segments of the population. They quite nicely illustrate the effects of the inhomogeneities in patient response and geometry. The results for the four different tumor doses show relatively smooth evolution, as would be expected given the logistic curve. They also can help determine the likely outcomes for different true differences in survival.

The gains in QALY as a function of reduction of uncertainty (Table 1) in different variables provides insight into likely sources of variability that could hamper a conclusive trial. The relatively large gains obtained with perfect information about utilities highlights one of the reasons that viewing clinical trials as decision problems is difficult. However, the same table indicates that better knowledge about dysphagia can have a large impact, especially when compared to the other complication, xerostomia.

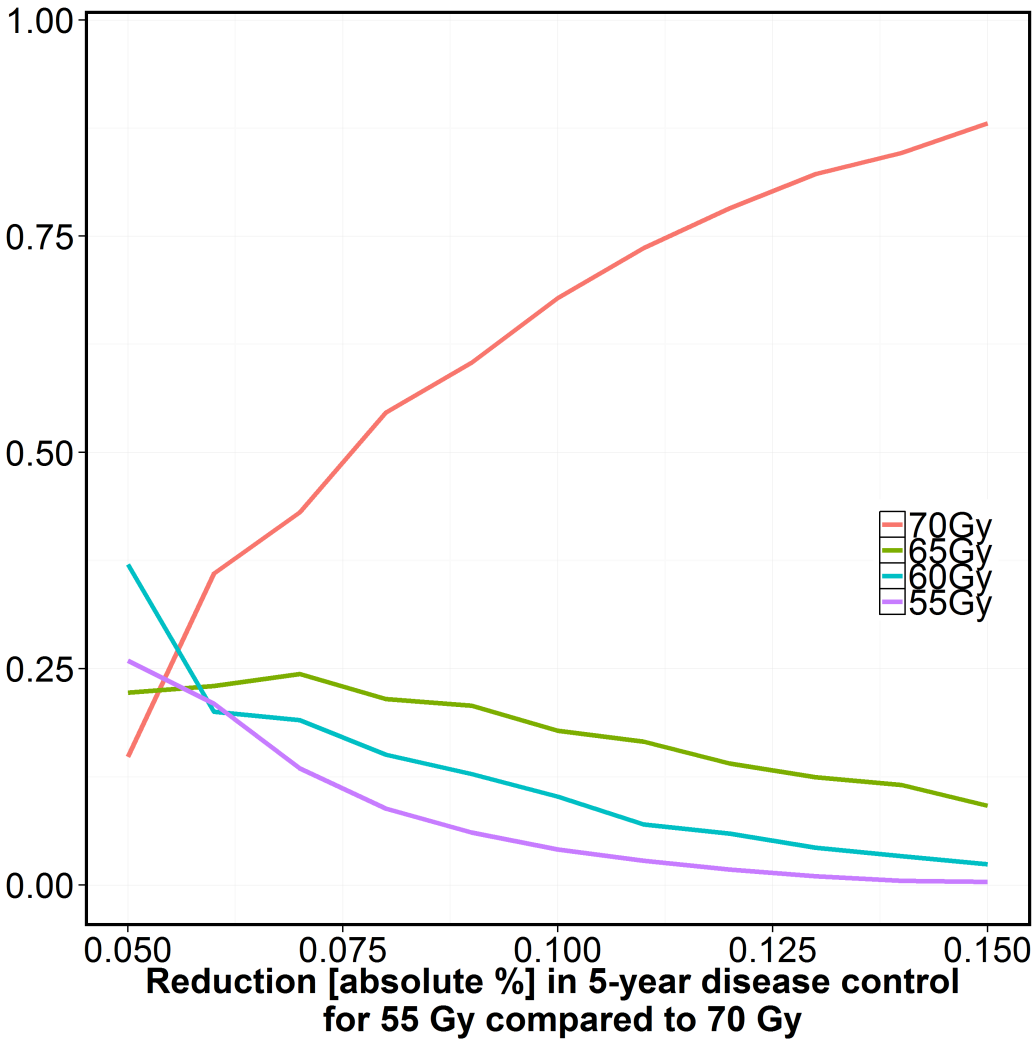

Figure 4: A plot of the fraction of patients for whom a given tumor dose (70, 65, 60, 55 Gy) would yield the most QALY's as a function of the difference in TCP(70 Gy) - TCP(55 Gy). The difference is relative to the TCP(70 Gy) reported in Ang *et al*[4].

This approach to modeling clinical trials has weaknesses. When compared to the typical inference approach (described above), our method requires much more quantitative work. To construct an adequate influence diagram, one needs reliable conditional probability tables for all the combinations of parent and child states. While some variables have binary states, some may be continuous variables leading to the imposition of a small number of ranges of values. In all of these cases, it is necessary to use realistic values. In addition, distributions among the various states of variables without parents must be determined. Although this allows an easy method for isolating different cohorts (for example, patients with a given T or nodal stage), one would need to measure, find in the literature, or estimate the distribution that will be present in the trial.

An important distinction between this method and the traditional method is the latter's reliance on null hypothesis significance testing (NHST). Under this approach, a trial outcome is considered "significant" if it obtains a p-value less than 0.05 (or 0.01 in more rigorous trials). However, there is numerous literature that describes in detail the flaws inherent in relying on such a method[19,20]. Our method explicitly models the outcomes as a function of effect size, thereby providing a much deeper understanding of what effect size is needed to make a study clinically relevant. This,

in turn, leads to a more accurate calculation of the power of a study. Overall, such an in-depth analysis in trial design will lead to less wasted effort and inconclusive trials.

Overall, modeling the clinical trial as we have provides a number of benefits which extend beyond the clinical trial itself. Such a model provides more insight into the likelihood that a given expected benefit or detriment will actually result in a clinically useful result. It also helps focus on the role that "nuisance" parameters play and the extent to which variability in them can reduce the significance of trial results. Thus, it may be prudent to either wait for better knowledge to be gained about a particular variable or to use that variable to distinguish between different arms of the trial. For example, dose de-escalation trials are more likely to result in significant results if patients whose tumors are near the swallowing muscles are separated from those whose are more distant, particularly if quality of life or dysphagia are to be measured as well.

The use of the simulation process has an added benefit outside of clinical trial design. As Fig. 4 points out, some fraction of patients are more likely to benefit than others from any given decision. Our model allows us to isolate the variables behind such differences, thereby providing clinical decision support.

We note that the distributions that were obtained from this model could be used to help design a clinical trial using Bayesian methods should one so desire. Prior probability distributions can be obtained from the model and incorporated into the design. In this particular case, a Bayesian approach is probably preferable over the standard Neyman-Pearson method. Knowledge of the actual difference in survival benefit, rather than a simple non-inferiority judgement is more likely to help clinicians. This is particularly true given the fact that such trials as these are unlikely to be repeated and reliance on simple p-values may result in unfortunate recommendations given the limited statistical sampling that will occur in a single trial[20].

Conclusion

A model for clinical trial design was described using probabilistic representation of the inhomogeneities within the patient population and the uncertainties regarding treatment response. The clinical trial was viewed as a decision problem and a cost function, in the form of a Markov model calculating quality adjusted life years, was implemented. The model was applied to a current issue in radiation oncology, namely the possible advantage of dose de-escalation in HPV+ oropharyngeal cancer. Using a probabilistic senstivity approach, results for different possible trial outcomes were calculated in the form of distribution of QALY's, the value of information, and the fraction of patients who would benefit from different strategies. This approach offers a number of advantages of traditional inference approaches, although it does require more effort up front. The overall goal of this method is to provide information regarding the critical variables so thagt clinical trials can be designed that will yield the most clinically useful information.

References

1. Speigelhalter DJ, Abrams KIR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. John Wiley and Sons, Chichester, England, pages 181-183, 2004.
2. Fiteni F, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials, J Visc Surg, 2014; 151: 17-22.
3. Müller P, Xu Y, Thall PF. Clinical trial design as a decision problem. Appl Stochastic Models Bus Ind 2016, 33: 296-301.
4. Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tan PF, Westra WH, Chung CH, Jordan RC, Lu C et al. Human papillomavirus and survival of patients with oropharyngeal cancer. New Engl J Med 2010, 353: 24-35.
5. Masterson L, Moualed D, Liu ZW, Howard JE, Dwivedi RC, Tysome JR, et al. De-escalation treatment protocols for human papillomavirus-associated oropharyngeal squamous cell carcinoma: a systematic review and meta-analysis of current clinical trials. European journal of cancer, 2014, 50: 2636-2648.
6. Phillips M, Smith W, Parvathaneni U, Laramore G. The Role of PET in the treatment of occult disease in head and neck cancer: a Modeling Approach. Int J Radiat Oncol Biol Phys 2011, 79: 1089-1095, 2011.

7. Kjaefulff UB, Madsen AL. Bayesian networks and influence diagrams: a guide to construction and analysis. Springer-Verlag, New York, 2008.
8. Keeney RL, Raiffa H. Decisions with multiple objectives: Preferences and value tradeoffs. Cambridge University Press, Cambridge, 1993.
9. Smith WP, Doctor J, Meyer J, Kalet IJ, Phillips MH. A decision aid for IMRT plan selection in prostate cancer based on a prognostic Bayesian network and a Markov model. Artif Intell Med. 2009; 46(2): 119-130.
10. Briggs A. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics, 2000, 17: 479-500.
11. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group6. Medical Decision Making. 2012 Sep;32(5):722-32.
12. Lyman JT, Wolbarst AB. Optimization of radiation therapy: III. A method of assessing complication probabilities from dose-volume histograms. Int J Radiat Oncol Biol Phys. 1987; 13: 103-109.
13. Miah AB, Gulliford SL, Clark CH, Bhide SA, Zaidi SH, Newbold KL, Harrington KJ, Nutting CM. Doseresponse analysis of parotid gland function: What is the best measure of xerostomia?. Radiotherapy and Oncology. 2013 Mar 31;106(3):341-5.
14. Christianen ME, Schilstra C, Beetz I, Muijs CT, Chouvalova O, Burlage FR, Doornaert P, Koken PW, Leemans CR, Rinkel RN, de Bruijn MJ. Predictive modelling for swallowing dysfunction after primary (chemo) radiation: results of a prospective observational study. Radiotherapy and Oncology. 2012 Oct 31;105(1):107-14.
15. van der Laan HP, van de Water TA, van Herpt HE, Christianen ME, Bijl HP, Korevaar EW, Rasch CR, van T Veld AA, van der Schaaf A, Schilstra C, Langendijk JA. The potential of intensity-modulated proton radiotherapy to reduce swallowing dysfunction in the treatment of head and neck cancer: A planning comparative study. Acta oncologica. 2013 Apr 1;52(3):561-9.
16. Meregaglia M, Cairns J. A systematic literature review of health state utility values in head and neck cancer. Health and Quality of Life Outcomes. 2017 Sep 2;15(1):174.
17. Ramaekers BLT, Joore MA, Grutters JPC, van den Ende P, De Jong J, Houben R, Lambin P, Christianen M, Beetz I, Pijls-Johannesma M, et al. The impact of late treatment-toxicity on generic health-related quality of life in head and neck cancer patients after radiotherapy. Oral Oncology; 2011; 47: 768-774.
18. Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research. Pharmacoeconomics. 2006; 24(11):1055-68.
19. Servick K. It will be much harder to call new findings 'significant' if this team gets its way. Science, Jul 25, 2017. doi:10.1126/science.aan7154.
20. Nuzzo R. Scientific method: statistical errors. Nature. 2014; 506: 150-152.