APENDIX A – Optimization objectives

A.1. Target constraints and considerations

For the CTV_{70} the same minimum constraint will be used as clinical standard, in order to prevent under-dosage to the conventional CTV70, i.e. D_{99} of 66.5Gy, with a D_{max}<107 before the boost/adapt plan. There will be no maximum dose with consequences to the CTV80-CTV70, but the optimization will include an objective to reduce the dose to 70 Gy as much as possible. For the CTV80 similar to the conventional constrained for CTVs 99% of the volume should receive 95% of the prescribed dose, which depends on the phase of the study.

A.2. Organ are dose optimization constraints

Per clinical standard the following critical organs at risk have the following radiation dose constraints:

Organ at risk	**Type of constraint**	**constraint**
Spinal cord	Maximum dose	<54.25Gy
Brainstem	Maximum dose	< 63.0 Gy
Optic nerve, chiasmus and retina	Maximum dose	< 60.0 Gy
Cochlea	Maximum dose	< 52.5Gy
Mucosal rim	Maximum dose	< 74Gy
Mandible	Maximum dose	< 74Gy

Deviations from prescriptions dose CTV_{adapt}, CTV_{70} and CTV_{57} are permitted to meet the above criteria.

All other organ at risks are objectives with desired to dose as low as feasible, but no consequences are linked to the doses they are planned to receive. In addition, the mucosa and mandibular will be excluded from the boost dose, i.e. have a constrained of <74Gy.

APPENDIX B - Study Schedule

B.1. Treatment Evaluations

The table below summarizes the pretreatment (initial screening, baseline measures prior to enrollment), on treatment, post-treatment phases of the trial, and the evaluations that would be required during these phases.

Table 4. Summary of Study Evaluations			
Screening	Baseline	Weekly treatment visits	Follow Up
Physical examination	X	X	X
CBC with differential	X		Check for unresolved >Grade 3 AE's
Basic Metabolic Panel	X	X	X
Pregnancy test (for female at reproductive age)	X		
Performance status evaluation	X	X	X
Contrast CT head and neck	X		X
18FDG PET/CT	X	Week 4	As indicated
PRO surveys	X	Weekly	6, 12, 24 months
CT-MRI verification scan	X	X	2-3 months
Alcohol/Tobacco urine test	**X***	**Week 1 and 3**	**8-12 wks, 6 , and 12 months***
Hearing/Audiology	X		8-12 wks, 12mo and 24 mo +/- 4 wks
Sialometry/ Trismus	X	Week 3 and 6	Every 1 years
Medical Photography of tumor/mucositis	X	Week 3 and 6	8-12 wks, 6, 12, 24 months
MBS +/-2 wks	X		3-6, 18-24, 60 months
Video-Stroboscopic examination	X		6, 24 months +/-4wks
Nutrition	X	X	

*** Indicates research related procedures. All other procedures are standard of care**

A complete history and physical to include performance status, recent weight loss, percent of weight loss, usual weight, and concurrent non-malignant disease and its therapy must be recorded. This information is to be collected during the clinic visit as per standard of care. Patient-reported smoking status prior to treatment (current, former, never, and pack years) will be collected.

A standard CT scan of the head and neck with IV contrast, MRI and a PET/CT scan are required workup components that should be done within 4 weeks prior to starting treatment. Study related contrast enhanced CT, MRI and PET/CT scans will be acquired in week 3 to delineate the tumor volume mid-treatment and determine the CTV_{adapt}.

Patients will be evaluated independently by the head and neck surgeon, the medical oncologist, and the radiation oncologist per institutional guidelines.

Pretreatment symptom scores and Quality of life (QOL) will be assessed for all patients upon enrollment into the study. These will be assessed as Physician-Reported as well as Patient-Reported Outcomes (PRO) surveys as per

institutional standard of care practice. Physician Reported Toxicity grading uses the CTCAEv5.0. The Patient Reported Outcomes are symptom measures assessed using the FACT-HNSI-10, the MD Anderson Dysphagia Inventory (MDADI), the MD Anderson Symptom Inventory (MDASI), theHead and Neck Xerostomia Questionnaire (XeQOLS), and Health Questionnaire (EQ-5D-3L). These surveys will be administered to the patient by the clinical staff prior to starting therapy as per standard of care.

Swallowing evaluation will be performed by videofluoroscopic examination of swallowing (i.e., modified barium swallow [MBS] study) prior to any treatment and with administration of the Performance Status Scale-Head and Neck (PSS-HN) and MD Anderson Dysphagia Inventory (MDADI) questionnaire. Functional assessment will record presence of tracheostomy, or feeding tube.

For the functional outcomes component, patients will have 3 brief additional tests performed at the same intervals as modified barium swallow studies. Collectively, these functional measures should take no more than 10 minutes to acquire, and add no radiation exposure or discomfort to clinical procedures for the modified barium swallow study.

B.2. Toxicity and symptom assessment

Evaluation at baseline, during and after treatment is summarized in Table 4. Treatment related symptom scores and quality of life (QOL) will be assessed each week during weekly scheduled treatment visits with the radiation oncologist as well as during each follow up specified times after completion of therapy. Like baseline measures, these will be assessed using both physician and patient reported outcomes (PROs). These data are collected by the clinical or research staff and inputted online through the web portal managed by the Biostatistics department. These can also be administered by paper or electronically using a computer terminal or computer tablet/handheld device, or by verbal/telephone interview.

The M.D. Anderson Symptom Inventory for Head and Neck Cancer (MDASI-HN) is a patient-reported outcome questionnaire designed to measure severity or burden of systemic and HNC-specific symptoms and their interference with or effect on patients' daily functioning [30,31]. This 28-item multi-symptom inventory measures includes 13 core items ("systemic symptoms": pain, fatigue, sleep, etc.), 9 head and neck-specific items ("local symptoms": dry mouth, mucus, shortness of breath, taste, etc.), and 6 interference items (activity, work, relations, etc.). The core MDASI items have been validated for use in cancer patient populations regardless of the specific diagnosis or type of therapy [31] and thus can be used to compare overall burden of disease between different types of cancer. The 9 head and neck-specific items were validated internally with regard to construct and concurrent validity in a cohort of 205 HNC patients [30]. Internal consistency reliability is high in the core MDASI score, the 9 head and neck-specific items, and the 6 interference items (Cronbach's alphas of 0.72-0.92).

The M.D. Anderson Dysphagia Inventory (MDADI) is a written questionnaire to evaluate dysphagia-specific QOL in head and neck cancer patients [32]. Dysphagia (swallowing difficulty) is a top priority of HNC survivors and a driver of QOL after treatment [33,34]. The 20-item MDADI questionnaire quantifies an individual's global, physical,

emotional, and functional perceptions of his or her swallowing ability. In an internal validation in 100 patients with HNC, concurrent validity was found to be moderate by comparison with the Performance Status Scale (Spearman correlation, 0.47-0.61). Correlation with the physical functional subscale (Spearman correlation, 0.40) and emotional subscale of SF-36 (Spearman correlation, 0.36) demonstrated convergent and divergent validity, respectively, of the MDADI. Test-retest reliability (physical, 0.86; emotional, 0.88; functional, 0.88) and international consistency reliability (overall Cronbach's alpha, 0.96) were sound.

Performance Status Scale for Head and Neck Cancer Patients (PSS-H&N) [35] is a clinician-rated 3-item instrument rated by semi-structured interview consisting of 3 questions: normalcy of diet, public eating, and understandability of speech. The PSS-HN has been psychometrically validated and recommended by the National Comprehensive Cancer Network for measurement of swallowing and speech performance in patients with head and neck cancer.

Modified Barium Swallow (MBS) studies (aka videofluoroscopic swallowing studies) are routine, gold-standard radiographic procedures that objectively examine oropharyngeal swallowing function. During a videofluoroscopic study, bolus trials of radiopaque-labeled bolus volumes are swallowed by the patient in a series of different liquid and solid food preparations to identify oropharyngeal swallow physiology, aspiration, and pharyngeal residue.

B.3. Measurement of Response/Progression Local or Regional Relapse

Relapse is defined as reappearance of tumor after complete response. If possible, relapse will be confirmed by biopsy. Progression is defined as an estimated increase in the size of the tumor (product of the perpendicular diameters of the two largest dimensions) of greater than 25%, taking as reference the smallest value of all previous measurements, or the appearance of new areas of malignancy. Distant Metastasis are defined when clear evidence of distant metastasis (brain, bone, lung, liver, etc.); biopsy when possible.

APPENDIX C - Statistical and Analytical Plans

This is a safety and feasibility study to assess image guided hybrid hyper-fractionated dose escalation with mucosal sparing proton therapy in 18 patients. Feasibility is defined as at least 80% of the patients receiving the MTD complete treatment. Safety is defined that none of the patients that receive the MTD develop severe unacceptable toxicities (grade 4 mucositis, dermatitis, aspiration, myelopathy, osteonecrosis) in 6 months following radiotherapy (primary endpoint) and if rates of any grade 3 toxicity do not exceed 80% of the patients in 3 to 6 months following treatment.

Summary statistics, including mean, standard deviation and 95% confidence interval will be used to describe safety data. Toxicity will be tabulated and summarized by grade and type. For the secondary objective, RISE rate will be summarized by frequency, standard deviation and 95% confidence interval.

Descriptive statistics, including mean, standard deviation and 95% confidence interval, will be used to summarize the exploratory endpoints, such as videofluoroscopy, imaging changes, MDADI, QOL, blood sample analysis, tumor tissue analysis, Physician-Reported and Patient-Reported Outcomes (PRO). These exploratory endpoints will be used for hypothesis generation.

C.1. Statistical Hypotheses

The hypotheses is that image guided hybrid hyper-fractionated dose escalation with mucosal sparing proton therapy is a feasible and safe treatment for Locally Advanced HNC patients.

C.2. Analysis of Datasets

Several large clinical trials have reported toxicity incidence rates of LAHNC after chemo-radiation, either with cisplatin, cetuximab or carboplatin [38–48]. This literature shows that acute grade 4 are very rare (~ 1-3% depending on the side effects) a summary of the reports data is depicted below.

Standard of care radiotherapy routinely demonstrates acute (i.e. <90 days post-therapy) Grade 3 toxicity rates are nearly ubiquitous (almost 90% for mucosal squamous carcinomas stage II-IV). Consequently, a 0.75 MTD rate at 3 months represents a conservative assessment of toxicity which is beginning to resolve still for many patients at 90 days/3 months post-chemo-radiation.

As mentioned before, Madani et al. [49] conducted a photon based (non-hyperfractionated or mucosal sparing) dose escalation study with tumor doses going up to a dose level I of 80.9Gy for 7 patients, subsequently followed by a dose level II of 85.9Gy for 14 patients. One patient (14%) in the dose level I cohort presented with a mucosal ulcer (mucositis grade 4) for 1 patients, which healed over time, and 5 (36%) in dose level II. The median recovery time of the 4 out of total 6 mucosal ulcers that healed was 4 months. The other 2 had follow-up times of 16 months. One patient (patient 4) presented with a mucosal ulcer at 8 months and died 2 months later due to its consequences.

C.3. Description of Statistical Methods

We will employ the time-to-event Bayesian optimal interval (TITE-BOIN) design [25] to find the maximum tolerated dose (MTD). Unlike the majority of existing phase I designs, which require suspending the accrual after treating each cohort of patients, the TITE-BOIN design allows for real-time dose assignment decisions for new patients while some enrolled patients' toxicity data are still pending. This shortens the trial duration and reduces the logistic difficulties caused by repeatedly suspending accrual. The TITE-BOIN works by predicting the dose-limiting toxicity (DLT) outcome for patients whose DLT data are pending based on their follow-up time. It is implemented in a simple way similar to the traditional 3+3 design, but is more flexible and possesses superior operating characteristics that are comparable to those of the more complex model-based designs, such as the time-to-event continual reassessment method (TITE-CRM).

We anticipate a monthly accrual rate of 2-3 patients.

For the purpose of dose finding, the DLT is defined as the toxicity occurred in the window of 3-6 months after the treatment as follows:

Mucositis	Grade=3
Dermatitis	Grade =3
Dysphagia	Grade ≥3
Hearing imp.	Grade ≥3
Xerostomia	Grade =3
Osteonecrosis	Grade =3
Pain	Grade =3
myelopathy	Grade =3
Aspiration	Grade =3

The target toxicity rate for the MTD is $\varphi = 0.75$ and the maximum sample size is 18. We will enroll and treat patients in cohorts of size 3. The DLT assessment window is T from 3 months to 6 months. To guide dose-escalation decisions, if the predicted DLT rate of the current dose cohort is ≤ 0.678, the next cohort of patients will be treated at the next higher dose level; if it is ≥ 0.803, the next cohort of patients will be treated at the next lower dose level. These boundaries were created when minimizing decision errors such that $\phi 1 = 0.6$ is the highest toxicity probability that is considered sub-therapeutic (underdosing) and $\phi 2 = 0.85$ is the lowest toxicity probability that is deemed overly toxic (overdosing). For the purpose of overdose control, doses will be eliminated from further examination if $Pr(pj > 0.75 \mid data) > 0.95$. The trial design is illustrated in Figure 1 and described through the following three steps:

1. Patients in the first cohort are treated at dose level 2.
2. To assign a dose to the next cohort of patients, count the number of patients ("No. pts"), the number of patients who experienced DLT ("No. DLTs"), and the number of pending patients ("No. pending") and their standardized total follow-up time ("STFT") at the current dose, and then make the dose escalation/de-escalation decision according to the rule displayed in Table 1. The STFT is defined as
3. STFT=sum of the standardized follow-up time for pending patients at the current dose / 3 months,

4. where the standardized follow-up time for a pending patient = the follow up time for a pending patient – 3 months. Here, subtracting 3 months from the follow-up time is used to account for the fact that the dose finding is based on the toxicity observed between 3-6 months. Here, subtracting 3 months from the follow-up time is used to account for the fact that the dose finding is based on the toxicity observed between 3-6 months. The STFT will be calculated using the online tool available at www.trialdesign.org, based on the follow-up time of pending patients

When using Table 1, please note the following:

   a. "Y&Elim" means de-escalating to the next lower dose and eliminating the current and higher doses from the trial to prevent treating any future patients at these doses because they are overly toxic. When the lowest dose is eliminated, stop the trial for safety. In this case, no dose should be selected as the MTD.
   b. If the current dose is the lowest dose and the decision table indicates dose de-escalation but no elimination, treat the new patients at the lowest dose.
   c. If the current dose is the highest dose and the rule indicates dose escalation, treat the new patients at the highest dose.
   d. For patient safety, if at the current dose, more than 50% of the patients' DLT outcomes are pending, suspend the accrual to wait for more data to become available. This rule corresponds to "Suspend" in Table 1.

In addition, at any time, if grade 4 mucositis, dermatitis, aspiration that do not resolve to a grade<3 in 3 months, CTCAEv5 grade≤3 myelopathy, and/or osteonecrosis or a radiation attributed grade 5 is observed in any patient, we immediately de-escalate the dose to the lower level.

3. Repeat 2 until the maximum sample size of 18 is reached or stop the trial if the number of patients treated at the current dose reaches 12.

Table 1: Decision table

No. pts	No. DLTs	No. pending	Escalate	STFT Stay	De-esc	No. pts	No. DLTs	No. pending	Escalate	STFT Stay	De-esc
3	0	<=1	Y			9	5	3	>=2.67	<2.67	
3	0	>=2		Suspend		9	5	4	>=3.87	<3.87	
3	1	<=1	Y			9	6	0	Y		
3	1	>=2		Suspend		9	6	1	>=0.96	<0.96	
3	2	0	Y			9	6	2	>=1.97	<1.97	
3	2	1	>=0.99	<0.99		9	6	3	>=2.99	<2.99	
3	3	0			Y&Elim	9	7	0		Y	
6	0	<=3	Y			9	7	1		>0.95	<=0.95

Table 1: Decision table

No. pts	No. DLTs	No. pending	Escalate	STFT Stay	De-esc	No. pts	No. DLTs	No. pending	Escalate	STFT Stay	De-esc
6	0	>=4		Suspend		9	7	2		>1.98	<=1.98
6	1	<=3	Y			9	8	<=1			Y
6	1	>=4		Suspend		9	9	0			Y&Elim
6	2	<=2	Y			12	0	<=6	Y		
6	2	3	>=1.58	<1.58		12	0	>=7		Suspend	
6	2	>=4		Suspend		12	1	<=6	Y		
6	3	0	Y			12	1	>=7		Suspend	
6	3	1	>=0.17	<0.17		12	2	<=6	Y		
6	3	2	>=1.49	<1.49		12	2	>=7		Suspend	
6	3	3	>=2.80	<2.80		12	3	<=5	Y		
6	4	0	Y			12	3	6	>=0.48	<0.48	
6	4	1	>=0.97	<0.97		12	3	>=7		Suspend	
6	4	2	>=1.99	<1.99		12	4	<=4	Y		
6	5	<=1			Y	12	4	5	>=1.57	<1.57	
6	6	0			Y&Elim	12	4	6	>=3.52	<3.52	
9	0	<=4	Y			12	4	>=7		Suspend	
9	0	>=5		Suspend		12	5	<=2	Y		
9	1	<=4	Y			12	5	3	>=0.30	<0.30	
9	1	>=5		Suspend		12	5	4	>=1.88	<1.88	
9	2	<=4	Y			12	5	5	>=3.47	<3.47	
9	2	>=5		Suspend		12	5	6	>=5.05	<5.05	
9	3	<=3	Y			12	5	>=7		Suspend	
9	3	4	>=1.59	<1.59		12	6	<=1	Y		
9	3	>=5		Suspend		12	6	2	>=0.45	<0.45	
9	4	<=1	Y			12	6	3	>=1.78	<1.78	
9	4	2	>=0.26	<0.26		12	6	4	>=3.12	<3.12	
9	4	3	>=1.74	<1.74		12	6	5	>=4.46	<4.46	
9	4	4	>=3.22	<3.22		12	6	6	>=5.79	<5.79	
9	4	>=5		Suspend		12	7	0	Y		
9	5	0	Y			12	7	1	>=0.29	<0.29	

Table 1: Decision table

No. pts	No. DLTs	No. pending	Escalate	STFT Stay	De-esc	No. pts	No. DLTs	No. pending	Escalate	STFT Stay	De-esc
9	5	1	>=0.26	<0.26		12	7	2	>=1.44	<1.44	
9	5	2	>=1.46	<1.46		12	7	3	>=2.60	<2.60	

Table 1 continued

No. pts	No. DLTs	No. pending	Escalate	STFT Stay	De-esc	No. pts	No. DLTs	No. pending	Escalate	STFT Stay	De-esc
12	7	4	>=3.75	<3.75		12	9	0		Y	
12	7	5	>=4.90	<4.90		12	9	1		>0.82	<=0.82
12	8	0	Y			12	9	2		>1.89	<=1.89
12	8	1	>=0.94	<0.94		12	9	3		>2.96	<=2.96
12	8	2	>=1.96	<1.96		12	10, 11	<=2			Y
12	8	3	>=2.97	<2.97		12	12	0			Y&Elim
12	8	4	>=3.99	<3.99							

Note: “No. treated” is the total number of patients treated at the current dose level, “No. DLTs” is the number of patients who experienced DLT at the current dose level, “No. with data pending” denotes the number of patients whose DLT data are pending at the current dose level, “STFT” is the standardized total follow-up time for the patients with data pending, defined as the total follow-up time for the patients with data pending divided by the length of the DLT assessment window. “Y” represents “Yes”, and “Y&Elim” represents “Yes and Eliminate”. When a dose is eliminated, all higher doses should also be eliminated.

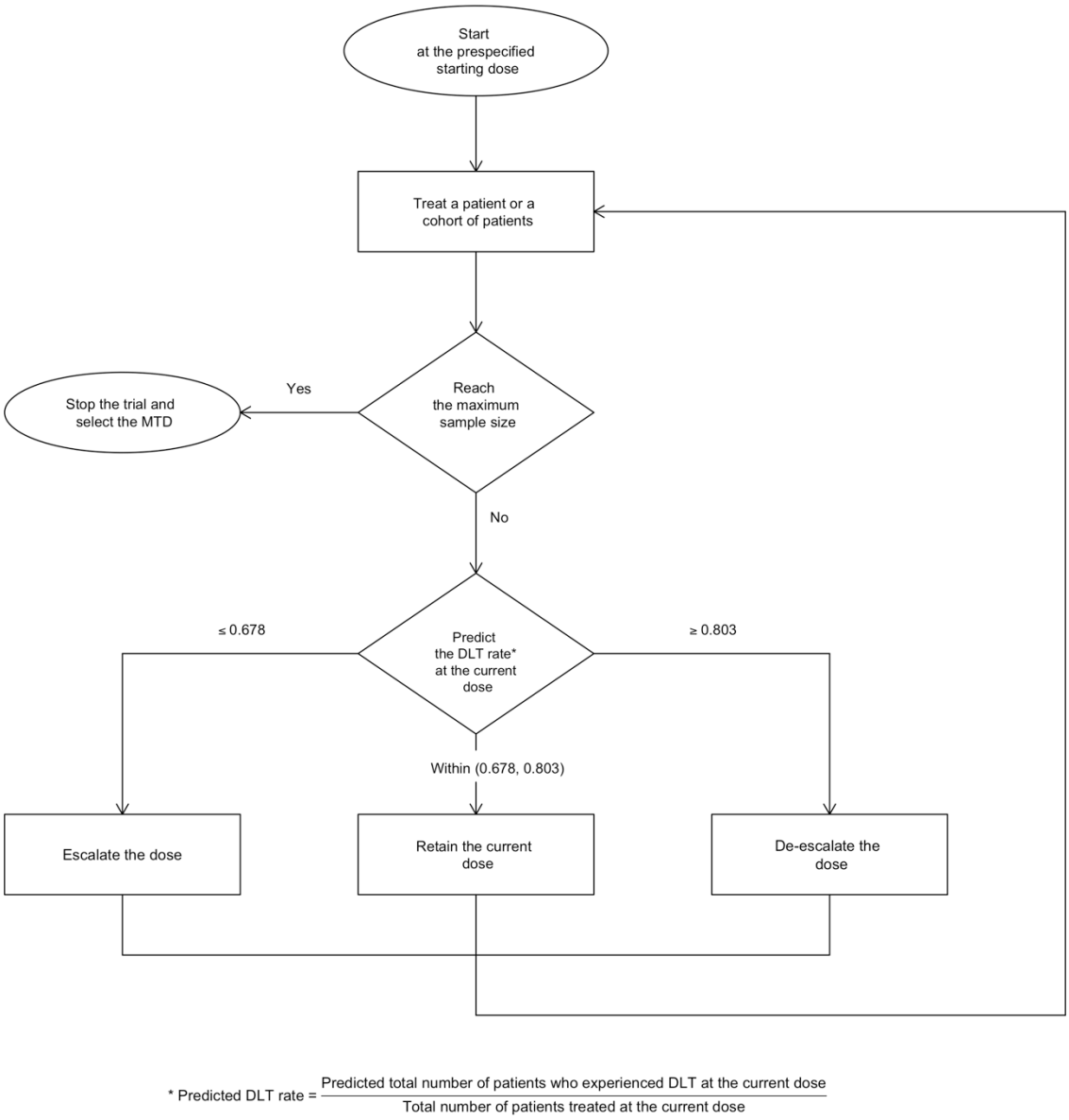

Figure 1. Flowchart for trial conduct using the TITE-BOIN design

After the trial is completed, select the MTD based on isotonic regression as specified in Yuan et al. (2018). This computation is implemented by the shiny app "TITE-BOIN" available at *http://www.trialdesign.org*. Specifically, select as the MTD the dose for which the isotonic estimate of the toxicity rate is closest to the target toxicity rate. If there are ties, select the higher dose level when the isotonic estimate is lower than the target toxicity rate and select the lower dose level when the isotonic estimate is greater than or equal to the target toxicity rate.

Data collection will be performed and entered into the requisite database for statistical analysis monthly by Head and Neck Radiotherapy Section research nursing staff. After every cohort of 3 patients are treated, we will provide the data to our statistical collaborators Lei Feng and Ying Yuan to determine the dose assignment for the next cohort of patients.

Operating Characteristics

Table 2 shows the operating characteristics of the trial design based on 1000 simulations of the trial using shiny app "TITE-BOIN" available at *http://www.trialdesign.org*. The time to toxicity is simulated from a Weibull distribution, with 50% of the DLTs occurring in the second half of the assessment window, and the patient accrual follows a Poisson process at the rate of 2 patients per month. The operating characteristics show that the design selects the true MTD, if any, with high probability and allocates more patients to the dose levels with the DLT rate closest to the target of 0.75.

Table 2: Operating characteristics

	Dose 1	Dose 2	Dose 3	# Patients	% Early Stopping
Scenario 1					
True DLT rate	0.56	0.75	0.9		
Selection %	11	72.2	15.8		1
# Pts treated	2.04	9.37	3.25	14.7	
Scenario 2					
True DLT rate	0.4	0.5	0.75		
Selection %	0.6	29.9	69.5		0
# Pts treated	0.3	6.69	8.27	15.3	
Scenario 3					
True DLT rate	0.75	0.95	0.98		
Selection %	57.1	28.8	0.2		13.9
# Pts treated	5.79	8.85	0.46	15.1	