ELSEVIER

Seminars in
RADIATION
ONCOLOGY

Comparing Intensity-Modulated Proton Therapy With Intensity-Modulated Photon Therapy for Oropharyngeal Cancer: The Journey From Clinical Trial Concept to Activation

Steven J. Frank, MD,[*] Pierre Blanchard, MD, PhD,[†] J. Jack Lee, PhD,[‡] Erich M. Sturgis, MD, MPH,[§] Merrill S. Kies, MD,[‖] Mitchell Machtay, MD,[¶] Bhadrasain Vikram, MD,[#] Adam S. Garden, MD,[*] David I. Rosenthal, MD,[*] G. Brandon Gunn, MD,[*] C. David Fuller, MD, PhD,[*] Katherine Hutcheson, PhD,[§] Stephen Lai, MD, PhD,[§] Paul M. Busse, MD, PhD,[**] Nancy Y. Lee, MD,[††] Alexander Lin, MD,[‡‡] and Robert L. Foote, MD[§§]

Intensity-modulated proton therapy minimizes the incidental irradiation of normal tissues in patients with head and neck cancer relative to intensity-modulated photon (x-ray) therapy and has been associated with lesser treatment-related toxicity and improved quality of life. A phase II/III randomized trial sponsored by the US National Cancer Institute is currently underway to compare deintensification treatment strategies with intensity-modulated proton therapy vs intensity-modulated photon (x-ray) therapy for patients with advanced-stage oropharyngeal tumors. After significant input from numerous stakeholders, the phase III portion of the randomized trial was redesigned as a noninferiority trial with progression-free survival as the primary endpoint. The process by which that redesign took place is described here.

[*]Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX.
[†]Department of Radiation Oncology, Gustave Roussy Cancer Campus, Villejuif, France.
[‡]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX.
[§]Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX.
[‖]Department of Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX.
[¶]Department of Radiation Oncology, UH Cleveland Medical Center, Cleveland, OH.
[#]Clinical Radiation Oncology Branch, National Cancer Institute, Bethesda, MD.
[**]Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA.
[††]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY.
[‡‡]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, PA.
[§§]Department of Radiation Oncology, The Mayo Clinic, Rochester, MN.

Supported in part by Grant CA016672 from the National Cancer Institute, US, National Institutes of Health to The University of Texas MD Anderson Cancer Center; Grant U19 CA021239 from the National Cancer Institute, US; and a Hitachi Grant Award.
Conflicts of interest: S.J. Frank: Advisory board member/consultant, Varian Medical Systems, funding from Hitachi Ltd.

*Address reprint requests to Steven J. Frank, MD, Department of Radiation Oncology, Unit 1422, The University of Texas MD Anderson Cancer Center, 1400 Pressler Street, Houston, TX 77030. E-mail: sjfrank@mdanderson.org

https://doi.org/10.1016/j.semradonc.2017.12.002

Introduction

Identifying the most appropriate primary endpoint may be the most important feature in the design of any clinical trial. The primary endpoint forms the basis or metric for assessing efficacy within study arms and for planned comparisons of efficacy between study arms. Once that endpoint is established, design considerations ensue in congruence with the various options regarding the phase of the clinical trial.[1] Ideally, the clinical trial design provides investigators with the ability to evaluate their primary objective in an unbiased manner, thus adding confidence to the acquired results.

The Department of Radiation Oncology at The University of Texas MD Anderson Cancer Center, under the aegis of a multi-institutional NIH/NCI-sponsored U19 cooperative agreement (2U19CA021239-35) with the Massachusetts General Hospital and IROC St. Louis, proposed a study to evaluate a potentially less toxic deintensification approach for delivering conformal radiation therapy to patients with cancer of the oropharynx. This approach involves the use of intensity-modulated proton therapy (IMPT), which is thought to reduce or eliminate the incidental irradiation of normal tissues associated with intensity-modulated [photon or x-ray] radiation therapy (IMRT) (Fig. 1).

Preliminary support for this concept came from dose distribution analyses that consistently showed superior dosimetry with IMPT for the treatment of head and neck cancers compared with IMRT[2-4] and from retrospective comparisons suggesting clinical benefits.[5-7] As for prospective evidence, the first 50 patients with oropharyngeal cancer (OPC) treated with IMPT experienced no grade 4 or 5 toxicity, and the 2-year actuarial rates of overall survival (OS) and progression-free survival (PFS) were 94.5% and 88.6%.[8] Also, a 1:2 case-matched control analysis of IMPT vs IMRT for OPC at MD Anderson revealed no significant differences in OS (hazard ratio (HR) = 0.55, 95% CI: 0.12-2.50, $P = 0.44$) or in PFS (HR = 1.02, 95% CI: 0.41-2.54, $P = 0.96$) between patients treated with IMPT vs IMRT. Third, a report of patient-reported outcomes (PROs) after IMPT vs IMRT for OPC,[9] obtained with the MD Anderson Symptom Inventory for Head and Neck Cancer (MDASI-HN) module, compared symptoms before treatment (baseline), during treatment (acute phase), within the first 3 months after treatment (subacute phase), and afterward (chronic phase). The 5 most common symptoms were found to be problems with food taste (mean score 4.91 on a 0-10 scale), dry mouth (4.49), swallowing or chewing (4.26), lack of appetite (4.08), and fatigue (4.00). Among the top 11 symptoms, changes in taste and appetite during the subacute and chronic phases favored the use of IMPT (all $P < 0.05$). During the subacute phase, the mean (±standard deviation) for the top 5 MDASI scores were 22% lower among patients who received IMPT (5.15 ± 2.66 for IMPT vs 6.58 ± 1.98 for IMRT, $P = 0.01$).

Despite this early evidence, irrefutable demonstration of the clinical superiority of proton therapy with level 1 evidence has yet to be accomplished. When the concept for the U19-supported clinical trial comparing IMPT with IMRT for OPC was developed, by consensus the main outcome of interest was the cumulative incidence of late-onset grade ≥3 treatment-related toxicity (scored according to the National Cancer Institute's Common Terminology Criteria for Adverse Events [CTCAE]) during the 2 years after completion of radiation

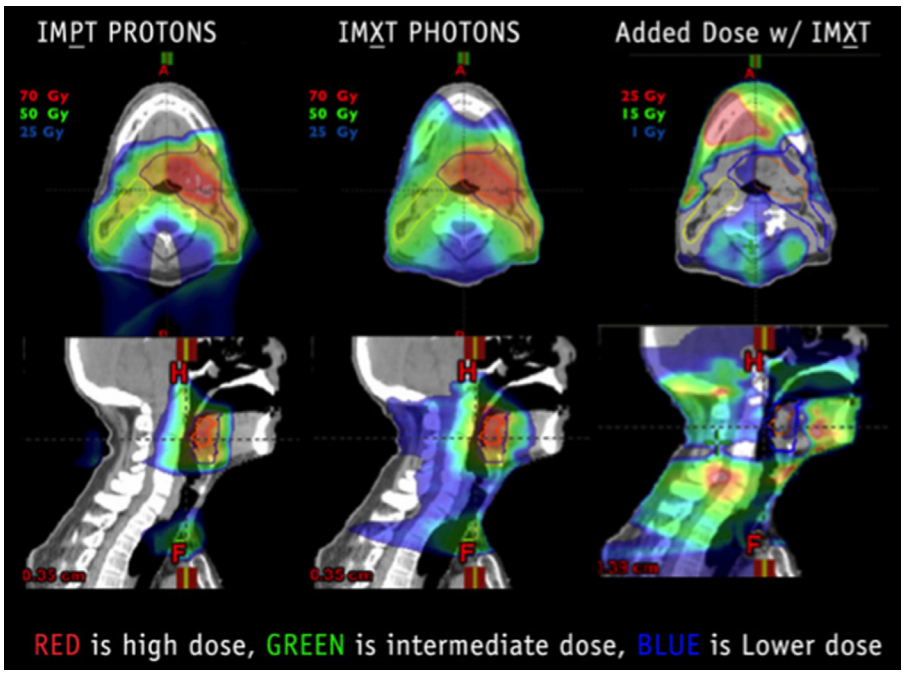

Figure 1 Axial (top) and sagittal (bottom) views of treatment plans used to assess dose distributions associated with intensity-modulated proton therapy (IMPT) (left) and intensity-modulated photon (x-ray) radiation therapy (IMXT) (middle). The images at right illustrate the additional dose associated with IMXT relative to IMPT. (Color version of figure is available online.)

therapy. In other words, the primary endpoint was a time-to-event endpoint, defined as the time from the start of the chronic period (defined as 90 days after completing radiation therapy) to the onset of a grade ≥3 treatment-related toxicity occurring within the 2-year evaluation window after the completion of radiation therapy.

To evaluate this primary endpoint in the two study groups, we planned a randomized phase II/III design according to Korn et al.[1] Patients were to be randomized in a 1:1 fashion to either IMRT or IMPT (Fig. 2). The randomization would be stratified by human papillomavirus/p16 status (positive or negative), smoking status (never vs former vs current and pack-years), and use of induction systemic therapy (yes or no). Initially, the phase II portion of the trial was to be launched only at MD Anderson, with the plan to add collaborating institutions as the trial progressed toward phase III. The intent was to transition to the phase III part of the trial through NRG Oncology, a nonprofit research organization formed to conduct clinical research in oncology and to broadly disseminate study results to inform clinical decision-making and health care policy (https://www.nrgoncology.org). However, the initially proposed primary endpoint (rate of grade ≥3 treatment-associated toxicity at 2 years) was met with resistance even after numerous discussions with NRG Oncology's oversight committee. The major point of contention was use of an endpoint based on the CTCAE scale for the phase III portion of the trial, because of a perceived lack of objectivity (owing to its dependence on physician reporting) and insufficient sensitivity to account for the numerous forms of toxicity experienced by a patient, which could thereby dilute potential differences between the 2 treatments. These criticisms prompted the investigative team to seek an alternative primary endpoint, which ultimately led to 2 major amendments to the trial design: first to redefine the primary endpoint and second to modify the trial design to accommodate recommendations from collaborators.

The following alternative endpoints were then explored: dependence on a feeding tube after treatment, which is one of the most common types of grade 3 toxicity associated with radiation therapy for head and neck cancer that affects patients' quality of life[10]; and grade ≥3 weight loss (ie, loss of >20% original body weight). Use of this composite endpoint was thought to limit the subjectivity of the decision to place a feeding tube, and severe weight loss could be used as a marker of patient malnutrition to capture patients who would have required a feeding tube but either declined or had not been offered one. Support for this proposed endpoint came from a retrospective case-matched analysis to detect differences between IMPT and IMRT in terms of this predefined composite endpoint (grade ≥3 weight loss or the presence of a feeding tube), which revealed odds ratios of 0.44 (95% CI: 0.19-1.0, $P = 0.05$) at 3 months after treatment and 0.23 (95% CI: 0.07-0.73, $P = 0.01$) at 1 year after treatment. Stated another way, patients with OPC treated with IMPT had reduced rates of feeding-tube dependence and malnutrition relative to patients treated with IMRT, without jeopardizing oncologic outcome.[6]

Despite the finding of a significant difference in this composite endpoint at 3 months and 1 year, the NRG Oncology reviewers maintained that this endpoint was still too subjective and that even positive results would not conclusively show the superiority of IMPT. One specific comment was that the use of a feeding tube was not a reliable, objective primary endpoint. To minimize this subjectivity, NRG Oncology instead recommended focusing on a very late time point for feeding-tube dependence, such as 1 year after treatment. However, such a low event rate would have required an unfeasibly large sample size even if the trial were conducted at multiple centers.

Another alternative suggested by NRG Oncology was to use a PRO measure as a primary endpoint. Although the importance of including PROs in clinical research is acknowledged,[11] and indeed collection of PROs was built into the phase II/III trial design (Fig. 2), a major challenge in using this approach was that an optimal clinical endpoint, with sufficient preliminary data, could not be agreed upon to compute a trial sample size. The only reference available in support of this approach was a retrospective unmatched comparison of patients treated with IMRT or IMPT at MD Anderson, for which relatively few

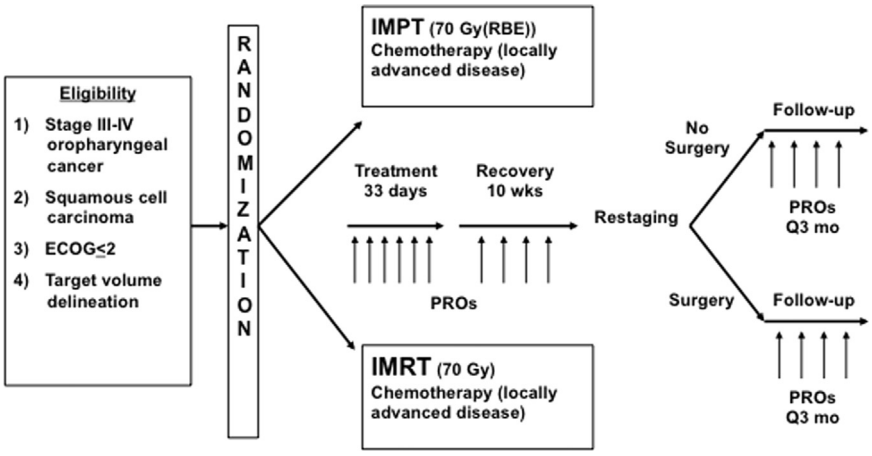

Figure 2 Treatment schema for the planned phase II/III trial of intensity-modulated proton therapy (IMPT) vs intensity-modulated (photon) radiation therapy (IMRT), both to a dose of 70 Gy, for patients with stage III-IV oropharyngeal cancer. PROs, patient-reported outcomes.

data and time points were available.[9] Although collecting PROs is strongly recommended, the debate continues regarding which PRO subsets—and which statistical analyses—should be used. To date, no randomized radiotherapy trial has used a PRO as its primary endpoint.[11] One exception is the ongoing Trans-Tasman Radiation Oncology Group (TROG) 12.01 trial (NCT01855451), which compares cisplatin with cetuximab-based chemoradiation for human papillomavirus-positive OPC and uses the MDASI-HN as its primary endpoint. However, in the absence of supporting data, TROG 12.01 is using the half-standard-deviation definition of minimum clinically important difference to compute the sample size for this study.

Waiting for the results of the phase II part of the study before embarking on the phase III part was another option, but concern was expressed that the accrual momentum would be lost during the 3- to 4-year delay between the end of the phase II trial and the start of the phase III portion (2 years of follow-up after the inclusion of the last patient + analysis + development of phase III protocol and regulatory submission). As an alternative, we eventually formed a OPC Patient Advisory Board and systematically considered patient preferences to support the choice and definition of an endpoint. Unfortunately, functional trade-offs or priorities regarding treatment outcomes have yet to be established for patients with OPC.[12] Focusing on a nononcologic outcome presented risks because (1) that outcome might not be the most relevant for every patient, (2) physicians may be biased toward the endpoint during the trial and could even subconsciously try to avoid it, and (3) if the trial findings were negative for that nononcologic endpoint, the magnitude of reduction with IMPT could still be relevant to the patient, or the trial could still be positive for other clinically relevant endpoints.

Collectively, these factors led to redesign of the phase III portion of the randomized trial as a noninferiority trial, with PFS as the primary endpoint, for the following reasons. First, PFS is a surrogate for OS.[13] Second, a precedent has been established for use of this design in RTOG 1016 (NCT01302834), a systemic therapy deintensification approach that compared cisplatin and cetuximab, given concurrently with IMRT, for p16-positive advanced OPC. Although cetuximab is more expensive than cisplatin, the increase in cost was considered to be offset by the reduction in toxicity. Third, this redesign, with PFS as the endpoint, ensures that tumor control is not jeopardized with the use of IMPT (ie, it ensures that the "efficacy" of IMPT is tested along with the toxic effects). Fourth, all of the analyses of toxicity and PROs planned in the initial phase II/III design will still be done; however, the current primary endpoint of the phase II portion (2-year cumulative incidence of grade ≥3 treatment-related toxicity) will now be reported as a secondary endpoint in the phase III trial. In that sense, the modified trial design allows robust characterization of all possible toxicity advantages (or, conversely, disadvantages) between IMPT and IMRT rather than arbitrarily focusing on a single toxicity deescalation endpoint that may be relevant to only a subset of patients. Finally, a cost-effectiveness analysis is planned in the revised phase III design to address the issue of the higher cost of IMPT.

Statistical Considerations for the Modified Design

Primary Objective and Justification of Sample Size

The primary outcome of this randomized phase III noninferiority trial, which has one planned interim analysis, is PFS rate at 3 years after treatment (Table). The 3-year PFS rate for the IMRT arm is assumed be 80% based on Ang et al,[14] preliminary data from RTOG 1016, and the MD Anderson experience with OPC.[15] A 9-percentage-point noninferiority margin will be used, similar to the one used in RTOG 1016. The corresponding HR is 1.535, based on the assumption that the time-to-event follows an exponential distribution.

Assuming a 1-sided type I error of 0.05 and an accrual rate of 10 patients per month, a sample size of 440 patients (220 randomized to each treatment arm) will yield 80% power to reject the null hypothesis corresponding to PFS (H_0: $\rho = 1.535$, where ρ is the HR [IMPT/IMRT]) and conclude that IMPT is noninferior to IMRT with respect to the specified 3-year PFS rate. The first interim analysis will be conducted after 72 of the expected 114 events have been observed. The method of Lan and DeMets,[16] with O'Brien-Fleming stopping boundaries,[17] will be used to stop early, if needed, for both noninferiority and superiority.

To have 440 evaluable patients, our sample size will be inflated to include an estimated 15% rate of patient loss owing to insurance-coverage denial of the treatment assigned by randomization. The final total sample size for the phase II/III randomized trial is 518 patients. This sample size calculation was done with East 6.3 (©2010, Cytel Inc., Cambridge, MA).

Secondary Objectives: Endpoints and Detectable Effect Sizes Given the Trial Sample Size

A total of 12 secondary endpoints for the phase III trial are planned, the first 3 of which focus on disease-related outcomes (patterns of failure, rates of OS and distant metastasis-free survival, and second primary cancers at 2 years); physician-graded toxicity (any CTCAE v4.0 grade ≥3 toxicity between 90 days and 2 years after completion of radiation, and a composite measure of weight loss and feeding tube placement); and PRO measures, assessed with the MDASI-HN, MD Anderson Dysphagia Inventory (MDADI), Functional Assessment of Cancer Therapy for patients with head and neck cancer (FACT-HN), and Xerostomia Questionnaire (XQ) (Table). Statistical considerations for the 2 endpoints related to toxicity are described later. In all cases, the detectable effect size, and the associated statistical method used for analysis, assume a power of 90% (with a type I [alpha] error of 0.01) and a 60% response rate for the PROs (which for 440 patients would be 263).

Physician-Rated Toxicity Outcomes

The first of the 2 physician-scored toxicity outcomes, grade ≥3 treatment-related toxicity occurring from 90 days to 2 years

Table Objectives for the Planned Phase III Trial Comparing Intensity-Modulated Proton Therapy With Intensity-Modulated (Photon) Radiation Therapy for Stage III-IV Oropharyngeal Cancer.

Primary objective

PFS rate at 3 y after treatment

Secondary objectives

1. Disease-related outcomes (patterns of failure, rates of overall and distant metastasis-free survival, and second primary cancers at 2 y)
2. Physician-graded toxicity
   A. Any grade ≥3 toxicity between 90 d and 2 y after completion of radiation
   B. Composite of >10% weight loss and presence of feeding tube at 2 y after treatment
3. Patient-reported outcomes
   A. MDASI-HN
   B. MDADI
   C. FACT-HN
   D. XQ
4. Quality-adjusted life-years (QALY) comparison (with EQ-5D)
5. Work productivity/impairment comparison (with WPAI)
6. Cost-benefit economic analysis
7. Correlational analyses of molecular profiles and OS or PFS
8. Correlative analyses of changes in serum biomarkers or HPV-specific cellular immune responses, measured at baseline and at 3 mo, with OS or PFS
9. Correlative analyses of banked peripheral blood samples and outcomes (samples collected at enrollment, at 6, 12, and 24 mo from the start of treatment, and at progression)
10. Correlative analyses of banked head and neck tissue samples to explore whether tissue-based markers can predict outcome
11. Banking peripheral blood and tissue samples for future interrogations
12. Physician-graded assessments of acute radiation-related side effects

Abbreviations: EQ-5D, the EuroQOL five-dimensional questionnaire on health-related quality of life; HPV, human papillomavirus; WPAI, Work Productivity and Activity Impairment.

after radiation, will be evaluated with a time-to-event analysis, and curves will be compared with 2-sided log-rank tests. In this way, we can detect a decrease in the cumulative incidence rate from 43% for IMRT to 29% for IMPT at 2 years (14% absolute reduction per 32.5% relative reduction, or HR = 0.635). This assumes an n of 440 patients, or approximately 243 events. The other outcome is a composite of ≥10% weight loss or the presence of a feeding tube, as suggested by Blanchard et al.[12] These outcomes will be measured at the end of the radiation treatment and at 6 months and 12 months thereafter. Findings will be evaluated by chi-square tests of equal proportions in the 2 treatment groups. We expect to be able to detect a difference in proportions from 75% to 58% (17% absolute reduction, or a 22.7% relative reduction).

Patient-Reported Outcomes

Six PRO questionnaires will be used for prospective data collection and assessment: the MDASI-HN, MDADI, FACT-HN, XQ, and 2 other surveys, the EuroQOL's 5-dimensional questionnaire on health-related quality of life (EQ-5D) and the Work Productivity and Activity Impairment (WPAI) questionnaire. Patients are asked to complete these questionnaires at baseline, once a week during treatment, at the end of treatment, every 3 months after treatment during the first year, every 4 months during year 2, and every 6 months up to 5 years.

Of primary interest will be changes in PRO from baseline to 3 months, as prior evidence suggests that most of the differences in PRO occur during this, the subacute phase.[9] For the MDASI-HN, we will evaluate mean scores derived from the top 5 items, and compare those mean scores between the 2 treatment groups during the acute, subacute, and chronic phases by using independent samples t tests at each phase, followed by a longitudinal analysis that includes area-under-the-curve analysis and a linear mixed-effects model. A power of 92% is associated with a standardized effect size of 0.50 (ie, detecting an effect of half of a standard deviation). In a second set of analyses, we will compare the proportions of patients in each treatment group with a clinically meaningful toxicity response (dichotomous outcomes; defined as an increase of half of a standard deviation from baseline being used to group patients as having [or not having] a response). This will be done with chi-square tests at the acute, subacute, and chronic phases, followed by repeated-measures logistic regression or a generalized linear mixed model. For these analyses, with an assumed proportion in 1 group of 50%, a minimum difference between 2 proportions of 0.23 (ie, a 46% relative reduction) can be detected with a 2-sided chi-square test. Similar analyses will be used for the MDADI, FACT-HN, and XQ data, with a mean composite score (MDADI), a mean total score (FACT-HN), or a summary score (XQ) derived from all test items and compared between treatment groups as described earlier. For these 3 surveys, the minimum detectable standardized effect size will be 0.478 (assuming an n of 264, with 2-sided independent samples t test, an alpha of 0.01, and 90% power).

Conclusions

We describe the process by which a phase III randomized trial was designed to test whether IMPT was truly a less toxic

deintensification strategy for patients with advanced OPC relative to the state-of-the art technique in photon therapy, IMRT. The redesign of the primary and secondary endpoints included input from numerous stakeholders across various disciplines and institutions, including patients with OPC. The revision of the clinical trial's phase III primary endpoint to PFS was approved by the UT MD Anderson Cancer Center IRB on September 15, 2017, with a planned date of phase III activation on December 15, 2017. Outcomes from this trial are expected to better define the value of proton therapy for patients with head and neck cancer.

Acknowledgment

The authors would like to acknowledge Thomas DeLaney for participation in discussions during the phase III development and review of the manuscript, and Christine Wogan for substantive editing of the manuscript.

References

1. Korn E, Freidlin B, Abrams J, et al: Design issues in randomized phase II/III trials. J Clin Oncol 30:667-671, 2012
2. Kandula S, Zhu X, Garden AS, et al: Spot-scanning beam proton therapy vs intensity-modulated radiation therapy for ipsilateral head and neck malignancies: A treatment planning comparison. Med Dosim 38: 390-394, 2013
3. Holliday EB, Kocak-Uzel E, Feng L, et al: Dosimetric advantages of intensity-modulated proton therapy for oropharyngeal cancer compared with intensity-modulated radiation: A case-matched control analysis. Med Dosim 41:189-194, 2016
4. Jakobi A, Bandurska-Luque A, Stützer K, et al: Identification of patient benefit from proton therapy for advanced head and neck cancer patients based on individual and subgroup normal tissue complication probability analysis. Int J Radiat Oncol Biol Phys 92:1165-1174, 2015
5. Romesser PB, Cahlon O, Scher E, et al: Proton beam radiation therapy results in significantly reduced toxicity compared with intensity-modulated radiation therapy for head and neck tumors that require ipsilateral radiation. Radiother Oncol 118:286-292, 2016
6. Blanchard P, Garden AS, Gunn GB, et al: Intensity-modulated proton beam therapy (IMPT) versus intensity-modulated photon therapy (IMRT) for patients with oropharynx cancer—A case matched analysis. Radiother Oncol 120:48-55, 2016
7. Holliday EB, Garden AS, Rosenthal DI, et al: Proton therapy reduces treatment-related toxicities for patients with nasopharyngeal cancer: A case-match control study of intensity-modulated proton therapy and intensity-modulated photon therapy. Int J Particle Ther 2:19-28, 2015
8. Gunn GB, Blanchard P, Garden AS, et al: Clinical outcomes and patterns of disease recurrence after intensity-modulated proton therapy for oropharyngeal squamous carcinoma. Int J Radiat Oncol Biol Phys 95:360-367, 2016
9. Sio TT, Lin HK, Shi Q, et al: Intensity-modulated proton therapy versus intensity-modulated photon radiation therapy for oropharyngeal cancer: First comparative results of patient-reported outcomes. Int J Radiat Oncol Biol Phys 95:1107-1114, 2016
10. Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, et al: Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. J Clin Oncol 26:3770-3776, 2008
11. Ringash J: Survivorship and quality of life in head and neck cancer. J Clin Oncol 33:3322-3327, 2015
12. Blanchard P, Volk RJ, Ringash J, et al: Assessing head and neck cancer patient preferences and expectations: A systematic review. Oral Oncol 62:44-53, 2016
13. Michiels S, Le Maître A, Buyse M, et al: Surrogate endpoints for overall survival in locally advanced head and neck cancer: Meta-analyses of individual patient data. Lancet Oncol 10:341-350, 2009
14. Ang KK, Harris J, Wheeler R, et al: Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 363:24-35, 2010
15. Garden AS, Dong L, Morrison WH, et al: Patterns of disease recurrence following treatment of oropharyngeal cancer with intensity modulated radiation therapy. Int J Radiat Oncol Biol Phys 85:941-947, 2013
16. Lan KKG, DeMets DL: Discrete sequential boundaries for clinical trials. Biometrika 70:659-663, 1983
17. O'Brien PC, Fleming TR: A multiple testing procedure for clinical trials. Biometrics 35:549-556, 1979