Quality of life in patients with oropharynx carcinomas: assessment after accelerated radiotherapy with or without chemotherapy versus radical surgery and postoperative radiotherapy

Allal, Abdelkarim Said; Nicoucar, Kevin; Mach, Nicolas; Dulguerov, Pavel

How to cite

ALLAL, Abdelkarim Said et al. Quality of life in patients with oropharynx carcinomas: assessment after accelerated radiotherapy with or without chemotherapy versus radical surgery and postoperative radiotherapy. In: Head & neck, 2003, vol. 25, n° 10, p. 833–839;discussion839–40. doi: 10.1002/hed.10302

This publication URL: https://archive-ouverte.unige.ch/unige:33121
Publication DOI: 10.1002/hed.10302

QUALITY OF LIFE IN PATIENTS WITH OROPHARYNX CARCINOMAS: ASSESSMENT AFTER ACCELERATED RADIOTHERAPY WITH OR WITHOUT CHEMOTHERAPY VERSUS RADICAL SURGERY AND POSTOPERATIVE RADIOTHERAPY

Abdelkarim S. Allal, MD,[1] Kevin Nicoucar, MD,[2] Nicolas Mach, MD,[3] Pavel Dulguerov, MD[2]

[1] Division of Radiation Oncology, University Hospital of Geneva, 1211 Geneva 14, Switzerland. E-mail: Abdelkarim.Allal@hcuge.ch
[2] Division of Head and Neck Surgery, Geneva, University Hospital, 1211 Geneva 14, Switzerland
[3] Division of Oncology, Geneva, University Hospital, 1211 Geneva 14, Switzerland

Accepted 20 December 2002
Published online 15 May 2003 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hed.10302

Abstract: *Background.* In oropharyngeal carcinomas, it is assumed that the effectiveness of the different treatment approaches is roughly equivalent, whereas the functional outcome after radical radiotherapy (RT) is superior to that associated with primary surgery. The aim of this study is to assess quality of life (QoL) outcomes of patients after two treatment strategies: radical surgery with postoperative RT and accelerated concomitant boost RT with or without chemotherapy.

Methods. Sixty patients who were disease free at least 1 year after treatment of oropharynx carcinoma were studied. Forty had been treated with radical RT (median tumor dose, 69.9 Gy in 5.5 weeks), and 20 had been treated with primary surgery and postoperative monofractionated RT (median dose, 60.2 Gy). Seven of the former patients received chemotherapy concomitantly with, and one before, RT. Functional outcome was assessed by the subjective Performance Status Scale for Head and Neck cancer (PSSHN) and the general QoL by the European Organization for Research and Treatment of Cancer Core QoL questionnaire (EORTC QLQ-C30). The unpaired *t* test was used to assess for significant differences between means.

Results. By use of the PSSHN module, scores were generally higher in the RT group, with a significant difference in the speech subscale ($p = .005$), a trend for a significant difference for the eating in public subscale ($p = .08$), and an insignificant difference for the normalcy of diet subscale ($p = .25$). When analyzed by tumor stage, no significant differences were observed for T1–2 tumors, whereas for patients with T3–4 tumors highly significant differences favoring the RT group became evident for all three subscales. Although no significant differences were observed using the EORTC QLQ C-30 functional scales, patients treated with primary surgery reported significantly more dyspnea (28 vs 12, $p = .04$) and appetite loss (30 vs 13, $p = .05$). In patients with T3–4 tumors, trends toward better scores favoring the RT group were observed for physical, role, emotional, and social functions, as well as a significantly better score for pain symptoms.

Conclusions. Although for early stages no clear advantage in QoL outcome was noted for the RT group compared with the surgery group, for advanced-stage disease an advantage favoring radical RT seemed apparent. For those patients, if an equivalency between the two treatment strategies could be assumed regarding oncologic results, then nonsurgical treatment should

Correspondence to: A. S. Allal

be considered the preferred option. © 2003 Wiley Periodicals, Inc. *Head Neck* **25:** 833–840, 2003

Keywords: accelerated radiotherapy; oropharynx cancers; quality of life; surgery

Depending on disease stage and institution preference, oropharyngeal carcinomas are treated with RT, surgery, or a combination of the two. Although no randomized trials have been carried out to help guide the choice of therapy, available data[1–4] suggest that the effectiveness of the different treatment approaches might be similar. Under these circumstances QoL considerations should play an important role in treatment choices. Despite a relative lack of formal QoL studies, functional outcome after radical RT is assumed to be superior to that associated with primary surgery, especially if postoperative RT is required in the latter case.[5] However, in recent years nonsurgical treatment approaches have become more aggressive, including the development of accelerated RT schedules and the concomitant administration of chemotherapy. The reported increase of severe late complications after some accelerated fractionation programs suggests that improved locoregional control might not be possible without greater chronic toxicity.[6,7] Documentation of QoL outcomes in patients treated with newer aggressive treatment approaches is thus of practical importance. At Geneva University Hospital, in the past patients with oropharynx carcinomas were treated preferentially by RT in early stages and surgery with postoperative RT in advanced stages. Recently, encouraging results obtained using concomitant boost RT, with or without chemotherapy,[8] led us to treat more and more patients with advanced disease in this fashion. This retrospective cross-sectional study compares QoL outcomes after accelerated RT, with or without chemotherapy, with those obtained after surgery and postoperative RT.

METHODS AND MATERIALS

Patient Characteristics. The study included patients treated between 1981 and 1998, who were disease free at least 1 year after treatment for carcinoma of the oropharynx. On the 84 patients treated with surgery in the period in question, 8 were lost to follow-up, and 27 were known to be alive (4 refused to participate, and 3 were with active disease). For the RT group, only patients treated with accelerated RT were considered. Of 93 patients treated, 44 were alive, 1 was lost to follow-up, and 3 had active disease. Patient participation was solicited during a routine or dedicated clinic visit. Sixty patients satisfying the inclusion criteria agreed to take part. Forty patients had been treated by radical RT with or without chemotherapy (RT group) and 20 by primary surgery with postoperative RT (surgery group). The median follow-up was 27 months (range, 12–82 months) for the RT group and 78 months (range, 16–200 months) for the surgery group. Patient characteristics are summarized in Table 1.

Treatment

RT Group. The RT schedule was a modified concomitant boost program that has been previously reported.[9] The schedule planned to deliver a total dose of 69.9 Gy in 41 fractions over a period of 38 days, using megavoltage beams. Involved sites and areas of potential microscopic disease (generally both sides of the neck down to the clavicles) received 50.4 Gy in 28 fractions over 5.5 weeks, and the boost to initial involved sites delivered 19.5 Gy in 13 fractions of 1.5 Gy given as a second daily fraction in a progressively accelerated fashion. The minimum interval between the two daily fractions was 6 hours. The median tumor dose was 69.9 Gy (range, 66.9–69.9 Gy).

In accordance with our treatment policy, seven patients had radical or selective neck dissection before RT, generally for bulky neck disease. Seven patients received chemotherapy concomitantly with RT, either alone ($n = 2$) or with neo-

Table 1. Patient characteristics.

Characteristics	RT group $n = 40$	Surgery group $n = 20$
Actual median age, y (range)	61 (42–83)	61 (48–75)
Gender: male/female	28/12	12/8
Tumor location (%)		
Tonsil/pillars	29 (72)	10 (50)
Base of tongue/vallecula	4	7
Soft palate/uvula	4	0
Overlapping	3	3
TNM classification (UICC 1992)		
T1–2 (%)	26 (65)	13 (65)
T3–4	14 (35)	7 (35)
N0–N1	26 (65)	10 (50)
N2–N3	14 (35)	10 (50)
TNM stage (UICC 1992)		
Stages I, II, III, IV	2, 12, 10, 16	2, 3, 4, 10

Abbreviations: RT, radiotherapy; UICC, International Union Against Cancer.

adjuvant and/or adjuvant chemotherapy ($n = 5$), and one patient received only neoadjuvant chemotherapy. Except for one patient treated with 5-fluorouracil and mitomycin C, all patients received cisplatin-based chemotherapy (associated with 5-fluorouracil in six patients). Six patients received three cycles, and two patients received two cycles.

Surgery Group. According to tumor location and extension, various surgical procedures were used. Surgery involving single subsites consisted of five partial pharyngectomies, four wide tonsillectomies, and one partial glossectomy. Multiple subsite surgery consisted of one of the aforementioned procedures and partial mandibulectomy in three patients, partial laryngectomy in four patients, total laryngectomy in one patient, and combined partial pharyngectomy and partial glossectomy in two patients. All but three patients also had radical ($n = 4$), modified radical ($n = 8$), or selective ($n = 5$) neck dissection. Neck dissection was bilateral in five patients. All patients received postoperative locoregional RT (median dose, 60.2 Gy; range, 45.6–68 Gy) with standard monofractionation. One patient also received three cycles of adjuvant cisplatin-based chemotherapy.

QoL Assessment. The assessment of QoL was performed by using two distinct questionnaires, one for disease-specific and one for general QoL aspects chosen for their proven validity and reliability: the Performance Status Scale for Head and Neck cancer (PSSHN) developed by List et al[10] and the European Organization for Research and Treatment of Cancer Core QoL questionnaire (EORTC QLQ-C30).[11]

1. PSSHN: This questionnaire is designed to assess dysfunction in the head and neck area. It is a clinician-rated tool consisting of three subscales assessing eating in public, understandability of speech, and normalcy of diet. The eating in public subscale assesses the impact of eating function disturbances on the social integration of the patient by reporting restrictions of settings and people present during food intake. The understandability of speech subscale rates the degree to which the interviewer is able to understand the patient's speech. The normalcy of diet subscale assesses the foods the patient is able to eat, with categories spanning the range from normal diet to no-oral feedings. The three subscales are rated from 0 to 100, with 100 representing normal function.
2. EORTC QLQ-C30: This is a patient self-rating questionnaire that is made up of six multi-item function scales measuring physical, role, social, emotional and cognitive functions, and overall QoL. Separate symptom scales are included to assess pain, fatigue, and emesis, and five single items are included to measure bowel function, breathing, appetite, and sleeping disturbances. A final item evaluates the economic consequences of the disease. All measurements are linearly transformed such that all scales range from 0 to 100, with higher scale scores representing a higher level of functioning for the six function multi-items and a higher level of symptoms/problems for the symptom/economic items.

Statistical Methods. The unpaired t test was used to assess for significant differences between means. A difference with a p value of .05 or less was considered significant.

RESULTS

PSSHN Scores. Table 2 displays the comparison for the three parameters for the RT and primary surgery groups. Globally, the scores were higher in the RT group, with a trend for a significant difference for the eating in public subscale scores ($p = .08$), a significant difference for speech subscale ($p = .005$), and an insignificant difference

Table 2. PSSHN function mean scores for the two groups.

	RT group $n = 40$ (SD)	Surgery group $n = 20$ (SD)	p value
Eating in public	84 (18)	73 (31)	0.08
Understandability of speech	95 (10)	81 (27)	0.005
Normalcy of diet	79 (19)	72 (27)	0.25

Abbreviations: RT, radiotherapy; SD, standard deviation.

Table 3. PSSHN function mean scores (SD) according to T stage.

	No. patients	Eating in public	Understandability of speech	Normalcy of diet
T1–2				
RT group	26	80 (20)	96 (9)	78 (21)
Surgery group	13	83 (24)	92 (12)	82 (22)
		$p = .7$	$p = .27$	$p = .56$
T3–4				
RT group	14	91 (12)	93 (12)	81 (18)
Surgery group	7	54 (36)	61 (35)	53 (25)
		$p = .002$	$p = .005$	$p = .008$

Abbreviations: RT, radiotherapy; SD, standard deviations.

for the normalcy of diet subscale. When splitting the results obtained in the two groups by tumor stage (T1–2 vs T3–4), in patients with T1–2 tumors the differences were insignificant for the three parameters, whereas for patients with T3–4 tumors all differences became highly significant (Table 3).

In the surgery group, patients with T1–2 tumors had significantly higher scores for the three subscales than did patients with T3–4 tumors. For surgically treated T1–2 and T3–4 patients, respectively, the eating in public subscale scores were 82.7 and 54 ($p = .04$), the understandability of speech subscale scores were 92 and 61 ($p = .007$), and the normalcy of diet subscale scores were 82 and 53 ($p = .01$). In contrast, no significant differences were observed between T1–2 and T3–4 patients in the RT group. For radically irradiated T1–2 and T3–4 patients, respectively, the eating in public subscale scores were 80 and 91 ($p = .06$), the understandability of speech subscale scores were 96 and 93 ($p = .33$), and the normalcy of diet subscale scores were 78 and 81 ($p = .69$). Moreover, no significant differences in mean scores were observed between patients treated with and without chemotherapy.

EORTC QLQ-C30 Scores. For the functional scales no significant differences were observed between the two groups, including the global QoL scale sores, which were similar (74 vs 69, $p = .4$) (Table 4). For the symptom scales no significant differences were noted, whereas for single items, patients treated with primary surgery reported significantly more dyspnea (28 vs 12, $p = .04$) and appetite loss (30 vs 13, $p = 0.05$). The results according to the T stage are summarized in Table 5. In patients with T1–2 tumors, global QoL score was similar (72 vs 71) in the two treatment groups, as were the other parameters except the social function score, which was significantly better in the surgery group (95 vs 78, $p = .03$); moreover, similar scores were observed in both groups in the functional scales as measured by the PSSHN module. In contrast, in patients with T3–4 tumors, a trend to significantly better scores favoring the RT group were observed for physical (86 vs 67, $p = .07$), role (90 vs 71, $p = .06$), emotional (83 vs 67, $p = .09$), and social (88 vs 64, $p = .07$) functions, as well as a significantly better score for pain symptoms (4 vs 21, $p = .008$) and trend to a better score for dyspnea (12 vs 33, $p = .08$).

DISCUSSION

In advanced-stage head and neck cancers, modifications in fractionation schedules and the con-

Table 4. EORTC QLQ-C30 mean scale scores for the two groups.

	RT group $n = 40$ (SD)	Surgery group $n = 20$ (SD)	p value
Functional scales			
Physical function	82 (22)	73 (26)	.16
Role function	84 (22)	89 (22)	.44
Emotional function	81 (24)	74 (19)	.3
Cognitive function	81 (23)	89 (14)	.16
Social function	81 (24)	84 (30)	.68
Global quality of life	74 (18)	69 (24)	.4
Symptom scales			
Fatigue	28 (26)	31 (31)	.77
Pain	15 (26)	22 (30)	.37
Nausea and vomiting	4 (15)	12 (27)	.1
Single items			
Dyspnea	12 (25)	28 (31)	.04
Sleep disturbance	23 (32)	20 (30)	.69
Appetite loss	13 (27)	30 (37)	.05
Diarrhea	4 (13)	3 (10)	.8
Constipation	13 (22)	10 (15)	.55
Financial impact	9 (23)	14 (28)	.47

Abbreviations: RT, radiotherapy; SD, standard deviation.

Table 5. EORTC QLQ-C30 mean scale scores for the two groups according to the T stage.

	T1–2 patients			T3–4 patients		
	RT group n = 26 (SD)	Surgery group n = 13 (SD)	p value	RT group n = 14 (SD)	Surgery group n = 7 (SD)	p value
Fuctional scales						
Physical function	80 (22)	75 (29)	.59	86 (20)	67 (20)	.07
Role function	80 (23)	83 (19)	.66	90 (18)	71 (25)	.06
Emotional function	79 (26)	78 (19)	.9	83 (22)	67 (19)	.09
Cognitive function	82 (19)	90 (14)	.21	80 (29)	88 (13)	.48
Social function	78 (26)	95 (14)	.03	88 (18)	64 (41)	.07
Global quality of life	72 (19)	71 (26)	.86	77 (16)	67 (20)	.19
Symptom scales						
Fatigue	32 (27)	31 (36)	.86	21 (24)	30 (23)	.39
Pain	21 (30)	22 (34)	.95	4 (7)	21 (21)	.008
Nausea and vomiting	6 (18)	14 (29)	.27	0 (00)	9 (25)	.16
Single items						
Dyspnea	13 (27)	26 (34)	.2	12 (25)	33 (27)	.08
Sleep disturbance	29 (33)	21 (32)	.42	12 (28)	19 (26)	.58
Appetite loss	14 (27)	31 (39)	.12	12 (28)	29 (36)	.25
Diarrhea	6 (16)	3 (9)	.43	0 (00)	5 (13)	.16
Constipation	12 (19)	5 (13)	.27	17 (28)	19 (18)	.84
Financial impact	10 (25)	14 (30)	.69	7 (19)	14 (26)	.48

Abbreviations: RT, radiotherapy; SD, standard deviation.

comitant administration of chemotherapy have both been shown to improve locoregional control compared with standard monofractionated RT alone.[7,12] However, in the absence of direct randomized comparisons, it remains uncertain whether the results of these new treatment strategies are equivalent to those obtained using radical surgery and selective postoperative RT. Taking into account the severe late complications reported after certain accelerated RT schedules[6] and the lack of prospective studies comparing functional outcomes after radical RT and primary surgery, documentation of QoL outcomes in patients treated with these new nonsurgical approaches has become a very pertinent area of clinical investigation. In this context, we compared QoL outcomes in patients with oropharyngeal carcinomas treated using accelerated RT with or without chemotherapy with those observed in patients treated with primary surgery and postoperative RT. We hypothesized that with the current aggressive RT schedule and the association of chemotherapy a decrease in functional outcome would be observed at least in some QoL parameters and/or subgroups of patients. For this purpose two validated questionnaires were used, one for disease-specific parameters (PSSNH) and one for general QoL measurements (EORTC QLQ-C30). Although the EORTC QLQ-C30 is a patient self-rating module, the PSSNH is clinician administered, with the disadvantage that the potential influence of the clinician on responses cannot be completely eliminated. Nonetheless, we retained this module because of the large experience with its use in head and neck cancer patients.

The overall results using the PSSNH module show a better functional outcome in the RT group (Table 2). Indeed, a trend to a significantly better mean score for eating in public and a very significantly better score for understandability of speech were noted in the RT group, whereas no significant difference was noted in the normalcy of diet subscale scores. Considering that advanced stages require larger tissue removal by surgery and could consequently negatively impact QoL outcomes,[13] we then studied T1–2 and T3–4 tumor patients separately. Interestingly, no difference was noted for the three subscale scores between the RT and surgery groups in T1–2 patients (Table 3). However, for T3–4 patients the differences were highly significant for the three subscale scores favoring the RT group. These results are partially in disagreement with the results reported by Harrison et al[5] on patients with base of tongue carcinomas using the same PSSHN module. Indeed, in the latter study, significant differences were noted even in case of T1–2 disease favoring patients treated with external RT and brachytherapy compared with patients treated with primary surgery. The difference is not due to lower scores in our RT group

patients (identical scores for the three subscales between the two series) but rather to lower scores noted in the surgery group in the series of Harrison et al. Moreover, this finding might reflect an effect of sample size in this latter study (five patients in the surgery group), as well as an effect of using a clinician-administered questionnaire as a self-rating one.

As expected, in the surgery group patients with T1–2 tumors had significantly higher scores for the three subscales compared with those of the T3–4 patients. This observation was already reported in other series ether using the PSSNH module[14] or the EORTC Head and Neck module.[15] In the RT group no significant differences were noted between the two T-stage groups. This result does not confirm those of the University of Florida series,[16] in which patients with higher stage disease had a higher degree of functional deficit. However, beside certain differences between the two series (different RT schedule and limitation of the current study to oropharynx tumors), in the Florida series no statistical comparison was done to substantiate the observed differences. Moreover, in our RT group no significant differences in mean scores were observed between patients treated with and without chemotherapy (data not shown).

The results obtained by using the EORTC QLQ-C30 showed no significant differences in the scores of the functional subscales between the RT and surgery groups. Interestingly, patients in the two groups reported similar global QoL scores despite some differences in single item/symptom scores. Indeed, patients treated with primary surgery reported significantly more dyspnea and appetite loss, as well as higher score for pain and nausea and vomiting. When splitting the results by T stage, in patients with T1–2 tumors, the global QoL score was similar in the two treatment groups, as well as the other parameters except the social function score, which was significantly better in the surgery group. There is no evident explanation for this significant difference, because the common RT side effects (eg, xerostomia) were observed in both groups; moreover, similar scores were observed in both groups in the functional scales as measured by the PSSHN module. In contrast, in patients with T3–4 tumors, a trend to significantly better scores favoring the RT group was observed for physical, role, emotional, and social functions, as well as a significantly better score for pain symptoms and trend to a better score for dyspnea. These results reflect the superior functional outcome in T3–4 oropharynx patients treated with radical RT as assessed by means of the PSSHN module.

As in the other studies addressing this specific question, our results should be interpreted with caution because of limitations inherent in the cross-sectional design of the study, namely, the difference in the length of follow-up and in sample sizes. Nevertheless, it seems very likely that differences in some QoL domains exist between patients treated with radical RT (with or without chemotherapy) and patients treated with radical surgery and postoperative RT, at least in some subgroups. This difference seemed most apparent in patients with advanced disease, in which functional outcome was better in the RT group, even considering the aggressive RT schedule and adjunctive concomitant chemotherapy. For patients with early disease, the two treatment approaches seemed equivalent in terms of functional outcome when considering all oropharynx subsites, although it is likely that for some subsites differences might exist, as suggested for base of tongue, in which better QoL outcomes have been reported after radical RT.[5]

Should we then move away from radical surgery to radical nonsurgical therapy for all advanced oropharyngeal cancers? Despite the ongoing movement in this direction, the definitive answer to this question awaits the conduct of studies confirming the oncologic equivalence of the two treatment strategies. In addition our results, and those of other recent QoL studies, require confirmation is required in prospective studies with larger sample sizes.

REFERENCES

1. Zelefsky MJ, Harrison LB, Armstrong JG. Long-term treatment results of postoperative radiation therapy for advanced stage oropharyngeal carcinoma. Cancer 1992; 70:2388–2395.
2. Amdur RJ, Parsons JT, Mendenhall WM, Million RR, Stringer SP, Cassisi NJ. Postoperative irradiation for squamous cell carcinoma of the head and neck: an analysis of treatment results and complications. Int J Radiat Oncol Biol Phys 1989;16:25–36.
3. Housset M, Baillet F, Dessard-Diana B, Martin D, Miglianico L. A retrospective study of three treatment techniques for T1-T2 base of tongue lesions: surgery plus postoperative radiation, external radiation plus interstitial implantation and external radiation alone. Int J Radiat Oncol Biol Phys 1987;13:511–516.
4. Wong CS, Ang KK, Fletcher GH, et al. Definitive radiotherapy for squamous cell carcinoma of the tonsillar fossa. Int J Radiat Oncol Biol Phys 1989;16:657–662.
5. Harrison LB, Zelefsky MJ, Armstrong JG, Carper E, Gaynor JJ, Sessions RB. Performance status after treatment for squamous cell cancer of the base of tongue–

a comparison of primary radiation therapy versus primary surgery. Int J Radiat Oncol Biol Phys 1994;30: 953–957.
6. Horiot JC, Bontemps P, van den Bogaert W, et al. Accelerated fractionation (AF) compared to conventional fractionation (CF) improves loco-regional control in the radiotherapy of advanced head and neck cancers: results of the EORTC 22851 randomized trial. Radiother Oncol 1997; 44:111–121.
7. Fu KK, Pajak TF, Trotti A, et al. A Radiation Therapy Oncology Group (RTOG) phase III randomized study to compare hyperfractionation and two variants of accelerated fractionation to standard fractionation radiotherapy for head and neck squamous cell carcinomas: first report of RTOG 9003. Int J Radiat Oncol Biol Phys 2000;48:7–16.
8. Allal AS, Bieri S, Miralbell R, et al. Combined concomitant boost radiotherapy and chemotherapy in stage III-IV head and neck carcinomas: a comparison of toxicity and treatment results with those observed after radiotherapy alone. Ann Oncol 1997;8:681–684.
9. Allal AS, de Pree C, Dulguerov P, Bieri S, Maire D, Kurtz JM. Avoidance of treatment interruption: an unrecognized benefit of accelerated radiotherapy in oropharyngeal carcinomas? Int J Radiat Oncol Biol Phys 1999;45: 41–45.
10. List MA, Ritter-Sterr C, Lansky SB. A performance status scale for head and neck cancer patients. Cancer 1990;66: 564–569.
11. Aaronson, NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365–376.
12. Calais G, Alfonsi M, Bardet E, et al. Randomized trial of radiation therapy versus concomitant chemotherapy and radiation therapy for advanced-stage oropharynx carcinoma. J Natl Cancer Inst 1999;91:2081–2086.
13. Colangelo LA, Logemann JA, Pauloski BR, Pelzer JR, Rademaker AW. T stage and functional outcome in oral and oropharyngeal cancer patients. Head Neck 1996;18: 259–268.
14. Zelefsky MJ, Gaynor J, Kraus D, Strong EW, Shah JP, Harrison LB. Long-term subjective functional outcome of surgery plus postoperative radiotherapy for advanced stage oral cavity and oropharyngeal carcinoma. Am J Surg 1996;171:258–261.
15. de Graeff A, de Leeuw JR, Ros WJ, Hordijk GJ, Blijham GH, Winnubst JA. A prospective study on quality of life of patients with cancer of the oral cavity or oropharynx treated with surgery with or without radiotherapy. Oral Oncol 1999;35:27–32.
16. Moore GJ, Parsons JT, Mendenhall WM. Quality of life outcomes after primary radiotherapy for squamous cell carcinoma of the base of tongue. Int J Radiat Oncol Biol Phys 1996;36:351–354.

EDITORIAL COMMENT

SOME CHALLENGES OF OBSERVATIONAL RESEARCH

Good clinical research is challenging. Realistically, many clinical research questions cannot be investigated experimentally with randomized controlled intervention trials. Thus, observational research methods, and all their limitations, are required. To conquer these challenges and draw valid conclusions, certain methodologic and analytical considerations are necessary.

Allal et al used a retrospective cohort study[1] to generate hypotheses on the very important clinical research question, "What is the best way to treat cancer patients?" Because survival differences between accelerated radiotherapy and surgery are not clear for their patient population, quality of life (QoL) and function might be the best outcomes to compare. Their study sheds light on several challenges of using an observational study design.

Biases. Biases compromise the process of seeking truth, and there are many forms. Many are insidious and unintentional and thus deserve attention. For example, selection bias produces incomparable groups and is one common form of bias. Investigators can provide reassurance by describing the method of sampling patients and by comparing baseline characteristics between groups. Blinding reduces certain biases by preventing subconscious influence over treatment procedures or outcomes assessment. There will always be uncontrolled biases, so it is appropriate to identify the most likely sources of bias openly and allow the reader to judge their importance.

Confounding Variables. Confounding variables distort the true relationship between two variables of interest. If not accounted, confounders might mislead an investigator into believing a false association. To illustrate, let us assume most early stage cancers were treated with surgery and most late stage cancers were treated with radiotherapy. Because early stage cancer has better survival, the surgery group would seem to fare better. This benefit, however, might be due solely to the difference in disease, not treatment. One can account for this confounding effect by stratifying on stage and recomparing treatment groups. Before stratifying for stage, Allal et al[1] found no advantage for radiotherapy. After stratifying, however, radiotherapy seemed advantageous. Comorbidity is another important confounder of QoL outcomes, because it might influence treatment choice and outcome. Age, gender, race, year of treatment, and many other variables potentially confound the fundamental relationship of interest. Sackett et al[2] highlight the importance of adjustment for confounders in their evidence grading scheme, in which an observational study that lacks adequate adjustment for

confounders is automatically relegated to a low evidence level. That said, some confounders might not be known, which is an inherent limitation of observational research.

Outcome Measures. The choice and method of outcome measurement is critical. QoL is an extremely important outcome, but its measurement is subjective and vulnerable to misinterpretation. Thus, reliability (reproducibility), validity (accuracy), and responsiveness (sensitivity to treatment effect) of QoL instruments should be tested before using them to compare outcomes. The clinically important difference, or at least clinical anchors of the score, is fundamental to QoL interpretation but is often lacking in head and neck cancer studies.[3] Comparisons of QoL outcomes are particularly vulnerable to misinterpretation when not accounting for baseline QoL. A high baseline QoL increases the chance of a high final QoL regardless of treatment effect. If there is an important difference in baseline QoL between treatment groups, it might confound the relationship between treatment and QoL outcome.

Multiple Testing. When one calculates statistical tests for multiple comparisons, some comparisons are likely to seem significant by chance alone. A p value < .05 simply means there is a 5% probability that the difference is not real, but rather by chance alone. If one tests 100 different comparisons, there is a high probability that some will be falsely positive. To highlight the "positive" findings from multiple comparisons is like bragging about the few shotgun pellets that actually hit the bull's eye. There are several ways to minimize spurious conclusions when performing multiple comparisons.

- Use a more rigorous threshold for statistical significance. For example, the Bonferroni correction uses a p value of .05 divided by the number of comparisons tested.
- Identify a priori the single one or two comparisons of primary interest. It is not valid to define the hypothesis after seeing the data, just as it is not legal to bet your horse after the race is over.
- Multiple secondary analyses can be considered exploratory. If no *a priori* hypothesis was defined, then the entire analysis is exploratory. Exploratory analyses generate hypotheses; they do not test hypotheses. Thus, one cannot draw conclusions from these analyses. The positive findings must be tested in a separate sample of patients. These exploratory analyses are analogous to identifying today's winning horse, so you can bet her on tomorrow's race.
- Consider the big picture and context. If all the results are consistent, explainable, and expected, it argues that the results are real. However, if there is a single positive finding or a positive finding that does not make sense, then it probably is falsely positive. Allal et al[1] found pain was significantly less in the advanced cancer subgroup compared with the early cancer subgroup of radiotherapy patients, which is unexpected and difficult to explain. This finding might be a false positive.

Statistical Power. For the unexpected negative results, it helps to quantify the probability that they are true negatives with a statistical power calculation. Statistical power increases with the size of the sample and the magnitude of difference in outcome between groups. Ideally, an investigator has a sense of the important difference he is looking for between groups and thus can calculate an estimate of the necessary sample size to achieve 80% power to detect that difference. Investigators often study as many subjects as they happen to have. Commonly, the sample is too small to discern a statistical difference in outcome between groups, even if the real difference is clinically important.

Despite these and many other difficult challenges with observational research, this type of research is necessary to expand our clinical knowledge. After all, this type of research has convinced people that smoking causes cancer. No experiment has randomly assigned people to smoking or not smoking to test the impact of smoking on human health.

REFERENCES

1. Allal AS, Nicoucar K, Mach N, Dulguerov P. Quality of life in patients with oropharynx carcinomas: assessment after accelerated radiotherapy with or without chemotherapy vs. radical surgery and postoperative radiotherapy. Head Neck. In press.
2. Phillips B, Ball C, Sackett D, et al. Levels of evidence and grades of recommendations. From: Centre for Evidence-Based Medicine. 2003 (http://cebm.jr2.ox.ac.uk, accessed January 9, 2003).
3. Schwartz S, Patrick DL, Yueh B. Quality-of-life outcomes in the evaluation of head and neck cancer treatments. Arch Otolaryngol Head Neck Surg 2001;127:673–678.