Published in final edited form as:

Cancer. 2017 January 01; 123(1): 12–16. doi:10.1002/cncr.30266.

The Role of Postoperative Chemoradiation for Oropharynx Carcinoma: A Critical Appraisal Revisited

Jay S Cooper, MD¹, Catherine Fortpied², Vincent Gregoire, MD³, Quynh-Thu Le, MD⁴, Thomas F. Pajak, PhD⁵, Qiang (Ed) Zhang, PhD⁶, and Jacques Bernier, MD⁷

¹Maimonides Medical Center, Brooklyn, NY ²EORTC Headquarters, Brussels, Belgium

³Université Catholique de Louvain, Brussels, Belgium ⁴Stanford University, Stanford, CA ⁵Retired RTOG ⁶NRG Oncology Statistics and Data Management Center, Philadelphia, PA ⁷Clinique de Genolier, Genolier, Switzerland

Precis

A recent review in this Journal, which criticizes NCCN's recommendations based, in part, on the RTOG 9501 and the EORTC 22931 trials, warrants comment. We herein provide additional data addressing many of the points and misinterpretations raised in that review which convincingly supports the use of concomitant chemotherapy and radiotherapy in patients with good performance status and high-risk pathologic features after surgery for mucosal head and neck squamous cell carcinoma.

Keywords

Postoperative; Chemotherapy-enhanced radiotherapy; Oropharynx; RTOG; EORTC

Sinha et al.'s review (1), which criticizes NCCN's recommendations based, in part, on the RTOG 9501 (2) and the EORTC 22931 (3) trials, warrants comment. We trust that the NCCN will also respond and therefore will limit our corrections to misunderstandings of the RTOG and EORTC trials and their respective manuscripts. Chemotherapy-enhanced radiotherapy (CERT) was recommended as standard therapy for post-operative patients based on the positive results in each of the original reports. There was no treatment recommendation in the combined paper (4).

The authors list 10 points that concern them:

1. **Lack of HPV status:** Both trials were designed in the early 1990's when there was no awareness of the coming importance of HPV. As far as we know, tumor HPV or p16 status was not assessed in any of the trials conducted during the period. Although HPV was not tested in RTOG 9501, the prevalence of HPV/p16-positive tumors is presumed to be low based on what we know from

Corresponding author: Jay S Cooper, MD, Maimonides Cancer Center, 6300 8th Avenue, Brooklyn, NY 11220, T. 718.765.2741, F. 718.765.2709, jcooper@maimonidesmed.org.

Conflicts of Interest: There are no conflicts of interest disclosures from any of the authors.

RTOG 9003, a trial that was conducted around the same time; only 39.5% of oropharyngeal carcinomas (OPC) in RTOG 9003 were p16-positive. In addition, since 42% of RTOG 9501 cases were OPC, assuming a 40% HPV-positive rate, it is estimated that only 17% of patients in this trial had HPV-positive tumors. In the EORTC 22931 trial, HPV/p16 status was also not assessed. Some data are nevertheless available at EORTC for the end of that period: using biological material from the EORTC study 24971, where a total of 358 patients were recruited between 1999 and 2002 (in Serbia, Czech Republic, Austria, United Kingdom, Germany, The Netherlands, Spain, Belgium, Hungary, France), positive p16 expression was detected by IHC in 16 of 119 patients with tumor biopsies (13%) and HPV DNA presence was detected by PCR and quantitative PCR (qPCR) in 20 of 79 patients with tumor biopsies (25%). Considering these data and the fact that the majority of the patients in EORTC 22931 were non-oropharyngeal sites (70%), the percentage of patients with HPV/p16-positive tumors in this trial is likely to be (very) low.

2. **Heterogeneous Study Population:** In the early 1990s, H&N squamous mucosal tumors were virtually always aggregated in national trials because they were considered to have similar (not identical) prognoses. The Gray's and the log-rank tests in RTOG 9501 reports were based on randomized, eligible patients for all time-related treatment comparisons with one exception. The primary endpoint, the rate of local-regional control (LRC), was also tested in all randomized patients (i.e., including ineligible patients) and the result was provided in the original report. No stratification was used with either test.

However, as can be seen from the multivariate analyses below, the results are consistent in both the original and long-term analyses. There are significant differences between the two arms for local-regional failure (LRF) and disease-free survival (DFS) in all initial analyses, even after adjusting for T-, N-classification, and tumor site. Although the difference did not reach statistical significance on long-term analysis, the hazard ratios were similar to the original estimates. Table 1.

3. **Failure to Report T and N classification:** The distribution of T and N classifications was not reported in the RTOG 9501 trial (it was provided in the original EORTC 22931 report), or in the pooled study, simply to abide by page limits imposed by the journals. However, the pT and pN categories for eligible patients in RTOG 9501 are well balanced as shown in Table 2.
4. **Exclusion after Randomization:** Consistent with the reporting of other national and cooperative group trials, the patient population used in the primary analysis, based on intent-to-treat, included only patients deemed eligible for the trial following review by the study chairs. This resulted in the exclusion of 9% (n=43) of patients in the initial RTOG report and 11% (n=49) in the subsequent report. However, as was stated in the original report "Including all ineligible randomized patients in the analysis did not change the result of the primary end point, which is LRC (HR=0.58, 95%CI 0.40–0.85,

$p=0.003$)". The same analysis with long term follow up shows that for LRC, HR=0.68, 95% CI 0.46–0.97, $p=0.02$. In addition, if the patients excluded from the long-term analysis were also excluded from the original report, the results for LRC for the original report would have been HR=0.64, 95% CI (0.43 to 0.95), $p=0.02$.

5. **Inconsistencies in numerical data:** Sinha et al. appear to have misunderstood the data. Their cited numbers (93 and 110) are from the figure of death as first DFS failure, not from that of overall survival, for which the corresponding numbers are 119 and 129. The difference between 252 and 248 is a consequence of the additional exclusions in the updated manuscript based on data that were submitted to RTOG HQ after the initial report. The 18% involved margin rate was based on data reported by the enrolling sites; however, central review corrected it to 10%, which was the number used in the joint analysis. There are no inconsistencies.
6. **Failure to report absolute effect sizes:** The absolute effects shown in Table 4 of Sinha et al. appear to have been derived by dividing the number of patients experiencing the event by the number of patients analyzed. Unfortunately, this approach neglects to adjust for different follow-up times. Instead, the effect size should be based on differences of survival rates at specified time points (for example, 2 and 5 years) estimated by the Kaplan-Meier method. For RTOG 9501, they are shown in Table 3.
7. **Use of unplanned subgroup analyses:** As we clearly stated in the combined paper (4), Sinha et al. are correct that the combined analysis was unplanned. However, our treatment recommendations were not based the combined analysis, but rather on the two antecedent original trials. Our combined analysis was done because failure to detect a statistically significant result does not exclude the possibility that one exists; we jointly examined the RTOG and EORTC trials hopefully to better understand why the RTOG trial did not detect a significant difference in OS while the EORTC did. In doing this analysis we discovered that the effect of CERT was essentially being driven by the sub-population common to both studies that had extra-capsular extension (ECE) and/or involved margins. This subgroup represented 64% of the combined randomized patients from the two trials: 59% of the RTOG 9501 trial and 70% of the EORTC 22931 trial. Of note, Figure 1 in the Sinha paper is visually misleading as the intersection of the two circles certainly does not represent 64% of the area of the two circles. As stated above, this subgroup analysis was an unplanned and exploratory, as was clearly stated in the joint manuscript (4) and the results have to be interpreted with appropriate caution. However, the forest plots presented in the paper show a rather homogeneous treatment effect across all subgroups within each study and across studies (within the subgroup of interest) and this suggests that the beneficial effect of CERT vs. RT in the subgroup is robust. In addition, the statistical powers for the observed hazard ratios in the combined analysis were quite robust: 95% for LRC, 89% for DFS and 85% for OS.

8. Failure to Perform Appropriate Multivariable Analyses: Please see point 2 above for RTOG 9501.

9a. No patient centered information reported: Sinha et al. discuss assessment of CERT-induced harm/toxicity (in their point 9 and associated Table 5) and we agree this is an important topic. More intensive regimens, by definition, are associated with more toxicity. All patients are not equally fit and we believe this is why NCCN publishes guidelines (not inflexible rules). We have never advocated treating *all* tumors that had ECE and/or involved margins with CERT (see point 10b below). However, for appropriately selected, fit patients, the choice of better tumor control at the price of increased acute toxicity (vs. decreased tumor control with less toxicity) is easy for most patients to make. Moreover, per the method of Doll (6), Sinha et al. appear to have miscalculated the number needed to harm (NNH) in their Table 5. The NNH for severe acute toxicity in should be 3 (not 2) and for severe acute and late toxicity the CERT data should read 161 of 206(78%), thus the NNH should be 4 (not 3).

Long-term reporting of RTOG 9501 (5):

9b. No adjuvant chemotherapy treatment effect in all eligible patients or in the OPC subgroup: Sinha et al. appear to confuse not detecting statistical significance with “no effect”. RTOG 9501 was designed with adequate power to detect differences in LRC within a two year time frame as specified in the original protocol. With longer follow up, the statistical power would decrease if the treatment effect lessens. For example, for LRC most of the events occurred within the first 2–3 years – a phenomenon typically observed in oncology trials. Based on the estimated effect size defined above for the oropharyngeal (OPC) and the Non-OPC subgroups, the treatment effect is consistent with long-term follow up but slightly smaller. As reported in Table 6 of the long-term paper, LRC and DFS are significantly different when the comparison is restricted to patients with ECE and/or involved margins with longer follow up, hence consistent with the combined analysis. Again we acknowledged the unplanned analyses in the long term paper, for example, for the primary endpoint LRC, we provided hazard ratios for OPC and Non-OPC groups and for the group with involved margins and/or ECE as an update. So the total of 18 unplanned analyses as mentioned in their discussion appears to be incorrect, as it mistakenly includes the planned analyses for all sites for the primary and secondary endpoints. Figures 1 and 2.

9c. Missing cause of death for study patients and unexplained high mortality in the CRT arm: Again, it is important to adjust for length of follow up when comparing endpoints such as time to death related to study cancer or not. Sinha et al.’s use of absolute percentages is not the best method to summarize these data because of this. With longer follow up, non-cancer-related deaths predominate, including from potential late toxicities of treatment. This is why we performed a cause-specific analysis for death from the study cancer, which demonstrated a significant risk reduction for the group with ECE and/or

involved margins. We have never advocated treating *all* tumors that had ECE and/or involved margins with CERT; clinical judgment based on co-morbidities and expected lifespans must prevail. However, we believe most “otherwise healthy” patients would rather risk longer-term death from a *potential* unexplained chemotherapy toxicity than short-term death from tumor.

- 9d. **Inconsistencies in numerical data:** Sinha et al. again appear to have incorrectly interpreted our data. Sixty-three patients were alive at the 10-year mark, but 11 died after 10 years; thus 52 were alive at the time of analysis. Also, 69 patients were censored (<10 years), for a total of 121 (52+69) patients, 87 of whom were free of disease at their last follow-up.
10. **Ineffectiveness of CRT on distant metastasis:** The primary hypothesis in RTOG 9501 was that concurrent high-dose cisplatin-based chemotherapy, when combined with radiation therapy, would improve LRC and the results of the study proved it. It would have been a serendipitous bonus also to observe a reduction in the occurrence of distant metastasis with concurrent chemotherapy; however, the rate of distant metastasis was not the primary endpoint of the trial and it was not statistically powered to test for a significant reduction in the rate of distant metastasis. To do so, a larger number of patients would have been needed because the distant metastasis rate is lower than the local-regional failure rate and the hypothesized chemotherapy effect on distant metastasis likely would be smaller.

We hope that this correspondence clarifies many of the points and misinterpretations raised by Sinha and colleagues. Level I evidence, as provided by RTOG 9501 and EORTC 22931, convincingly support the use of concomitant chemotherapy and radiotherapy in patients with good performance status and high-risk pathologic features after surgery for mucosal head and neck squamous cell carcinoma. There may well be patients who satisfy the eligibility criteria in either trial, but would not benefit from adjuvant CERT. The HPV infected patients could be such patients. However, our data does *not* suggest this, in fact, it suggests the opposite. We strongly encourage Sinha and his colleagues to perform level I randomized clinical trials to try to prove their points. Unproven conjectures without clinical trial validation will not advance our field or the care of our patients, even when offered as “critical appraisals”.

Acknowledgments

Funding: This project was supported by grants U10CA21661 (RTOG-Ops-Stat), U10CA180868 (NRG Oncology Operations) and U10CA180822 (NRG Oncology SDMC) from the National Cancer Institute (NCI).

References

1. Sinha P, Piccirillo JF, Kallogjeri D, Spitznagel EL, Haughey BH. The role of postoperative chemoradiation for oropharynx carcinoma: a critical appraisal of the published literature and National Comprehensive Cancer Network guidelines. *Cancer*. 2015 Jun 1; 121(11):1747–1754. [PubMed: 25588360]

2. Cooper JS, Pajak TF, Forastiere AA, Jacobs J, Campbell BH, Saxman SB, Kish JA, Kim HE, Cmelak AJ, Rotman M, Machtay M, Ensley JF, Chao KSC, Schultz CJ, Lee N, Fu KK. Postoperative Concurrent Radiation Therapy and Chemotherapy in High-Risk Squamous Cell Carcinoma of the Head & Neck: The RTOG 9501/Intergroup Phase III Trial. *N Engl J Med*. 2004 May 6; 350(19):1937–1944. [PubMed: 15128893]
3. Bernier J, Dumenil C, Ozsahin M, Matuszewska K, Lefebvre JL, Greiner RH, Giralt J, Maingon P, Rolland F, Bolla M, Cognetti F, Bourhis J, Kirkpatrick A, van Glabbeke M. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *N Engl J Med*. 2004 May 6; 350(19):1945–1952. [PubMed: 15128894]
4. Bernier J, Cooper JS, Pajak TF, van Glabbeke M, Bourhis J, Forastiere A, Ozsahin EM, Jacobs JR, Jassem J, Ang KK, Lefebvre JL. Defining risk levels in locally advanced head and neck cancers: A comparative analysis of concurrent postoperative radiation plus chemotherapy trials of the EORTC (#22931) and RTOG (# 9501). *Head Neck*. 2005; 27(10):843–850. [PubMed: 16161069]
5. Cooper JS, Zhang Q, Pajak TF, Forastiere AA, Jacobs J, Saxman SB, Kish JA, Kim HE, Cmelak AJ, Rotman M, Lustig R, Ensley JF, Thorstad W, Schultz CJ, Yom SS, Ang KK. Long-term Follow-up of the RTOG 9501/Intergroup Phase III Trial: Postoperative Concurrent Radiation Therapy and Chemotherapy in High-Risk Squamous Cell Carcinoma of the Head and Neck. *Int J Radiat Oncol Biol Phys*. 2012 Jun 30; 84(5):1198–1205. [PubMed: 22749632]
6. Doll HA. Making sense of the numbers: CER, EER, ARR, RR, RRR, NNT, CIs. *Found Years*. 2009; 5:1–7.

Figure 1.
Comparison of local-regional control and survival rates for all patients grouped by primary tumor site (oropharynx vs. non-oropharynx) and treatment (RT vs. CERT).

Figure 2.

Comparison of local-regional control and survival rates for patients who had extracapsular extension of disease and/or an involved surgical margin grouped by primary tumor site (oropharynx vs. non-oropharynx) and treatment (RT vs. CERT).

Table 1

Endpoints and Covariates for Locoregional Control, Disease-Free Survival, and Overall Survival From Initial and Long-Term Analyses

Endpoint	Report	Covariates	Hazard ratio (95%CI)	p-value
Local-regional control	Initial	None	0.61 (0.41–0.91)	0.011
		Stratifying variables	0.61 (0.41–0.90)	0.013
		Stratifying variable and site, T, N	0.59 (0.39–0.88)	0.009
	Long-term	None	0.73 (0.49–1.07)	0.098
		Stratifying variables	0.72 (0.49–1.06)	0.098
		Stratifying variable and site, T, N	0.69 (0.46–1.02)	0.063
Disease-free survival	Initial	None	0.78 (0.61–0.99)	0.038
		Stratifying variables	0.77 (0.61–0.98)	0.036
		Stratifying variable and site, T, N	0.79 (0.62–1.00)	0.052
	Long-term	None	0.88 (0.71–1.09)	0.247
		Stratifying variables	0.87 (0.70–1.09)	0.226
		Stratifying variable and site, T, N	0.88 (0.70–1.10)	0.255
Overall survival	Initial	None	0.84 (0.65–1.09)	0.190
		Stratifying variables	0.84 (0.64–1.09)	0.180
		Stratifying variable and site, T, N	0.85 (0.65–1.11)	0.237
	Long-term	None	0.89 (0.70–1.12)	0.308
		Stratifying variables	0.89 (0.70–1.12)	0.303
		Stratifying variable and site, T, N	0.90 (0.71–1.14)	0.392

Table 2

Pathologic Tumor and Lymph Node Classification for Eligible Patients in Radiation Therapy Oncology Group Trial 9501

pT	RT (n=210)	CERT (n=206)
T1	33 (15.7%)	21 (10.2%)
T2	49 (23.3%)	60 (29.1%)
T3	49 (23.3%)	58 (28.2%)
T4	79 (37.6%)	67 (32.5%)
pN	RT (n=210)	CERT (n=206)
N0	3 (1.4%)	4 (1.9%)
N1	10 (4.8%)	9 (4.4%)
N2a	10 (4.8%)	12 (5.8%)
N2b	137 (65.2%)	126 (61.2%)
N2c	45 (21.4%)	48 (23.3%)
N3	4 (1.9%)	7 (3.4%)
Unknown	1 (0.5%)	0

Table 3
Effect Sizes Based on Differences in Survival Rates at 2 and 5 Years From the Initial and Long-Term Analyses

Endpoint	Report	2 years		5 years	
		Estimates (%)	Absolute difference (%)	Estimates (%)	Absolute difference (%)
Local-regional control	Initial	72.4/81.5	9.1	68.4/80.6	12.2
	Long-term	74.3/81.5	7.2	71.8/80.1	8.3
Disease-free survival	Initial	44.8/53.9	9.1	25.4/35.0	9.6
	Long-term	45.9/55.3	9.4	29.9/38.5	8.6
Overall survival	Initial	57.1/64.1	7.0	37.3/44.6	7.3
	Long-term	58.7/64.7	6.0	37.1/45.8	8.7