INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS
www.redjournal.org

CLINICAL INVESTIGATION

Accuracy and Prognosis of Extranodal Extension on Radiologic Imaging in Human Papillomavirus-Mediated Oropharyngeal Cancer: A Head and Neck Cancer International Group (HNCIG) Real-world Study

Hisham Mehanna, PhD,[a] Ahmad K. Abou-Foul, MD,[a] Christina Henson, MD,[b] Caroline Kristunas, PhD,[a] Paul C. Nankivell, PhD,[a] Lachlan McDowell, MBBS,[c] C. René Leemans, PhD,[d] Michiel W.M. van den Brekel, PhD,[e] Christian von Buchwald, DMSc,[f] Kathrine Kronberg Jakobsen, MD,[f] Christian Grønhøj, MD, PhD,[f] Jacob Høygaard Rasmussen, PhD,[f] William M. Lydiatt, MD,[g] Vishal Gupta, MD,[h] Barton F. Branstetter, MD,[i] Jens Peter Klussmann, MD,[j] Barbara Wollenberg, PhD,[k] Benedikt Schmidl, MD,[k] Martina A. Broglie, MD,[l] Jan-Jaap Hendrickx, MD, PhD,[d] Daniel R. Awad, MS,[m] Robin Prestwich, PhD,[n] Anne-Laure Gaultier, PhD,[o] Marc Oliva, MD,[p] Sudhir Nair, MCh,[q] Anthony Noor, BDSc (Hons), MBBS,[r] Suren Krishnan, FRACS,[s] Zsuzsanna Iyizoba-Ebozue, MD,[n] Mantegh Sethi, MB BChir,[a] Irene H. Nauta, MD,[d] Daniel Zhao, PhD,[t] and Sue S. Yom, PhD[u,v]

[a]Institute of Head and Neck Studies and Education, University of Birmingham, Birmingham, United Kingdom; [b]Department of Radiation Oncology, Stephenson Cancer Centre, University of Oklahoma Health Sciences Centre, Oklahoma City, Oklahoma; [c]Department of Radiation Oncology, Princess Alexandra Hospital, Brisbane, Queensland, Australia; [d]Department of Otolaryngology-Head and Neck Surgery, Amsterdam University Medical Centre, Cancer Centre Amsterdam, Amsterdam, The Netherlands; [e]Department of Head and Neck Oncology and Surgery, Netherlands Cancer Institute - Antoni van Leeuwenhoek, Amsterdam, The Netherlands; [f]Department of Otorhinolaryngology, Head and Neck Surgery & Audiology, Copenhagen University Hospital - Rigshospitalet, Denmark; [g]Department of Surgery, Nebraska Methodist Hospital and Methodist Women's Hospital, Creighton University, Omaha, Nebraska;; [h]Department of Radiation Oncology, Mount Sinai Health System, New York, New York; [i]Department of Radiology, University of Pittsburgh, Pittsburgh, Pennsylvania; [j]Department of Otorhinolaryngology, Head and Neck Surgery, Medical Faculty, University Hospital of Cologne, Cologne, Germany; [k]Department of Otolaryngology, Head & Neck Surgery, Technical University Munich (MRI TUM), Munich, Germany; [l]Department of Otolaryngology-Head and Neck Surgery, University Clinic of the University Hospital Zurich, University of Zurich, Zurich Switzerland; [m]University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; [n]Department Clinical Oncology, Leeds Cancer Centre, Leeds, United Kingdom; [o]Service d'Imagerie, Hopital Européen Georges PompidouAPHP, Université Paris Cité, Paris, France; [p]Department of Medical Oncology, Head and Neck Cancer Unit/Drug Development Program, Institut Català d'Oncologia (ICO) L'Hospitalet, Barcelona, Spain; [q]Department of Head and Neck Oncology, Tata Memorial Centre, Mumbai, India; [r]School of Medicine, University of New South Wales, Sydney, New South Wales, Australia;

Corresponding authors: Hisham Mehanna, PhD; E-mail: h.mehanna@bham.ac.uk

Daniel Zhao, PhD performed the statistical analysis.

Disclosures: H.M. is the Director and a shareholder of Warwickshire Head and Neck Clinic; Chair of the Head and Neck Cancer International Group (HNCIG); and past president of the British Association of Head and Neck Oncologists. H.M. also reports receiving honoraria from AstraZeneca; Speakers Bureau on MSD, Sanofi Pasteur, Merck; Research funding from GSK Biologicals, MSD, Sanofi Pasteur, GSK Plc, AstraZeneca; travel accommodation expenses from Sanofi, Pasteur, MSD, Merck. W.L. declares he is the chair of American Joint Committee on Cancer (AJCC) 9th Version head and neck task force. M.O. has consulting/advisory arrangements with Merck, MSD and Transgene. Research support from Merck and Roche. The institution receives clinical trial support from Abbvie, Ayala Pharmaceutical, MSD, ALX Oncology, Debiopharm International, Merck, ISA Pharmaceuticals, Roche Pharmaceuticals, Boehringer Ingelheim, Seagen, Gilead. Travel accommodations expenses: BMS, MSD, Merck. S.Y. received research funding from Bristol-Myers Squibb and EMD Serono. C.H. receives payment as an ad hoc scientific advisory board member for EMD Serono. J.P.K. receiving honoraria from MSD, Merck and BMS; research Funding from MSD. All other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author, HM, on reasonable request.

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ijrobp.2025.04.003.

Int J Radiation Oncol Biol Phys, Vol. 123, No. 2, pp. 432–441, 2025

https://doi.org/10.1016/j.ijrobp.2025.04.003

[s]Department of Otolaryngology, Head & Neck Surgery, Royal Adelaide Hospital, South Australia, Australia;; [t]Department of Biostatistics and Epidemiology, Stephenson Cancer Centre, University of Oklahoma Health Sciences Centre, Oklahoma City, Oklahoma; [u]Department of Radiation Oncology, University of California, San Francisco, California; and [v]Department of Otolaryngology-Head and Neck Surgery, University of California, San Francisco, California

Received Nov 18, 2024; Revised Mar 6, 2025; Accepted for publication Apr 4, 2025

Purpose: Extranodal extension on radiology (iENE) is reported in single-center studies to be negatively prognostic in human papillomavirus-mediated oropharyngeal cancer (HPV + OPC) and is a major eligibility criterion for surgical treatment. However, studies report widely varying sensitivities, specificities, and interobserver correlation. In this research the prognostic power, sensitivity, and specificity of iENE in HPV + OPC in real-world practice are determined.

Methods and Materials: A retrospective cohort of 821 consecutive subjects with p16 + OPC, treated with surgery and/or chemoradiation therapy (CRT), from 13 multinational secondary hospitals in 9 countries between January 1, 1999 and December 31, 2020 was analyzed. The main outcomes were sensitivity, specificity, and overall survival (OS). Assessors were blinded to outcomes.

Results: Six hundred thirty-eight patients were included in the final analysis. A total of 109 of 394 (27.7%) with no iENE had ENE on histopathology (pENE), and 109 of 192 (56.8%) of patients with pENE were misclassified as having no iENE. iENE sensitivity and specificity were 44.5% (95% CI, 37.8%-51.4%) and 87.6% (95% CI, 84.1%-90.6%), respectively, and varied significantly between centers. Negative predictive value was 75.3% (95% CI, 72.3%-77.5%).

Subgroup analyses showed significantly increased sensitivity and specificity if patients had both computed tomography (CT) and magnetic resonance imaging (MRI): 84.6% (95% CI, 65.1%-95.6%, $P < .001$) and 94.5% (95% CI, 82.3%-99.4%, $P = .022$), respectively, compared with only CT or MRI alone. Specialist radiologists showed better specificities (89.14%; 95% CI, 85.69%-91.99% vs 46.67%; 95% CI, 21.27%-73.41%, $P < .001$) and similar sensitivities to nonspecialists.

On multivariable analysis, iENE positivity was *not* a statistically significant independent predictor of OS (adjusted hazards ratio [aHR], 1.50 [95% CI, 0.97-2.32; $P = .071$]) or disease-free survival (aHR, 1.41; 95% CI, 0.95-2.09; $P = .089$). Two proposals for amended TNM staging did not yield large improvements.

Conclusions: In current real-world practice, iENE showed widely varying and modest accuracy, and was not independently prognostic of outcomes in HPV + OPC. iENE accuracy and prognostic power increased significantly by using combined CT and MRI scanning, experienced head and neck radiologists and more inclusive diagnostic criteria. Validated consensus diagnostic criteria and protocols are urgently needed to enhance the clinical utility of iENE. Until then, clinicians should be cautious about making treatment decisions based on iENE.

Introduction

Over the past decade, transoral robotic surgery has increased in popularity, becoming the dominant treatment for early human papillomavirus-mediated oropharyngeal cancer (HPV + OPC) in many countries.[1] Consequently, extranodal extension detected on radiology (imaging-detected extranodal extension [iENE]) is increasingly being used to determine treatment selection for HPV + OPC.[2-4] Specifically, the absence of iENE is used as a main eligibility criterion for transoral surgery, with the aim of selecting those patients with no ENE on histopathology (pENE),[2,3] and so avoiding triple therapy (surgery and postoperative chemoradiation therapy [CRT]), with its considerable functional disability.

However, the evidence for the utility of iENE in HPV + OPC is not well established. Meta-analysis of data from mainly single centers, with small sample sizes, reported by one or a small team of radiologists, suggests that iENE carries prognostic importance in HPV + OPC, and indeed may be even more prognostic than pENE.[5] However, there was a large variation in the reported prognostic powers between the different studies. In addition, 2 systematic reviews report variable pooled sensitivities of 60% to 77% and specificities of 60% to 96%,[6,7] with widely varying and modest correlation coefficients between radiologists of 0.4 to 0.7,[6] with a pooled correlation coefficient of 0.72 (95% CI, 0.60-0.81).[7]

For iENE to be used reliably in routine clinical practice to determine treatment selection, it should demonstrate consistent and high sensitivities, specificities, negative predictive values, and prognostic power in a multicenter real-world setting, and not just in studies with centralized or highly controlled radiologic reporting that is not replicable in the routine practice. Here, to our knowledge, in one of the largest studies reported in the literature, we aimed to determine the accuracy and prognostic power of iENE in HPV + OPC in a multicenter, multinational real-world setting. To future proof findings, we also assessed 2 recent proposals for incorporating iENE into the American Joint Committee on Cancer TNM staging for HPV + OPC. Results are reported according to the STROBE guidelines.

Methods and Materials

Study design

This was a centralized individual patient data analysis on data from 14 head and neck cancer centers from 10 countries (Australia, Denmark, France, Germany, India, Spain, Switzerland, The Netherlands, United Kingdom, and United States). We included retrospective series and prospective cohorts with minimum cohort sizes of a total of 50 cases. Selected patients had to fulfill a priori inclusion criteria, including consecutive patients aged 18 years or more, diagnosed with a primary squamous cell carcinoma of the oropharynx between January 1, 1999 and December 31, 2020, and treated with curative intent by surgery, radiation therapy (RT), CRT, or a combination of these, and must have had computed tomography (CT) or/and magnetic resonance imaging (MRI) scans of the neck performed within 12 weeks before the start of treatment, and have data on clinical outcomes and follow-up (date of last follow-up if alive, date of recurrence or metastasis, and date and cause of death). Positivity on p16 immunohistochemistry was defined as strong nuclear and cytoplasmic staining in at least 70% of the tumor cells.[8] There were no limits on upper age or performance status. Radiological diagnosis of iENE for all cases from each center were reported by the same radiologist, and scored as present, absent or equivocal, using the criteria the radiologist used in their routine clinical practice, which included irregular nodal margins, extension into perinodal fat, matted or conglomerate nodes, and extension into other structures such as muscle or skin. To reflect real-world practice, the radiologists used the criteria that they normally use to report iENE in routine clinical practice. Patients with distant metastasis or recurrent disease at diagnosis were excluded. Potential selection bias was addressed through strict recruitment criteria, consecutively recruited cases, and a minimum size of cohort for inclusion, as well as blinding analysis from data collection. We addressed other forms of bias through sensitivity analysis.

Because this was an analysis of anonymized routinely collected clinical data, no approval was deemed necessary by the ethics committee of the lead site, the University of Birmingham, but each research site obtained any locally-required ethics and other approvals before start of data collection.

Outcomes

The primary outcome was overall survival (OS), defined as duration from end of treatment until death from any cause, or until last follow-up appointment, or until end of follow-up. The secondary outcomes were the sensitivity and specificity of iENE reported by the local radiologist compared with pENE diagnosed by the local pathologist (gold standard) in cases in which histopathology results were available because surgery was performed.

Statistical analysis

Analyses were preplanned—details of statistical analysis are provided in the Supplementary Materials section E1, Supplementary Methods. In brief, patient demographic and disease characteristics were compared between ENE+ and ENE− cases using the χ^2 test, Fishers exact test, t test, or Wilcoxon rank-sum test, as appropriate. Survival was compared between iENE+ and iENE− and between pENE+ and pENE− groups using the log-rank test and Kaplan−Meier survival curves were produced. Adjusted Cox proportional hazards models were fitted to estimate the effects of iENE and pENE on survival, adjusted by potential confounding factors, with log-likelihood ratio test P values used to assess the significance of each variable in the model. The sensitivity, specificity, positive predictive value and negative predictive value (and 95% CIs) of iENE in diagnosing pENE (gold standard) were calculated. Robustness to assumptions was assessed through sensitivity analyses. Subgroup analyses were also performed.

The 2 amendments to the TNM 8 classification system proposed by Huang et al,[9] were applied to the N stage and overall stage groupings, and Cox proportional hazards models were fitted to compare the OS of the groupings with the respective ones by TNM 8. Models were compared by their Harrell's C-indices.

The significance threshold was established initially at 0.05, with a Bonferroni's correction used in multiple comparison analyses. Analyses were performed using STATA/SE 17.0 (StataCorp. 2019. Stata Statistical Software: Release 16: StataCorp LLC).

Results

Patient demographic and disease characteristics

Overall, 821 subjects with HPV + OPC were included, treated between January 1, 1999 and December 31, 2020 at 13 centers in 9 countries, since the cohort from India had no HPV + OPC cases. There were 651 men and 178 women with a mean age of 59.9 years (SD, 9.40).

Performance status was available for only 442 of the 821 (53.8%) subjects, of whom 400 of 442 (90.5%) had performance status Eastern Cooperative Oncology Group (ECOG) Performance Status 0 or 1 and 42 of 442 (9.5%) status 2 or 3. A total of 260 of 821 (31.7%) had never smoked, 328 of 821 (40.0%) were previous smokers, and 174 of 821 (21.2%) were current smokers, with 59 of 821 (7.2%) of unknown status. A total of 539 of 821 (65.7%) had T1 and T2 tumors, 81 of 821 (9.90%) had TX or T0 tumors, and 153 of 821 (18.6%) had T3 and T4 tumors. A total of 169 of 821 (20.6%) had N0, 501 of 821 (61.0%) N1, 87 of 821 (10.6%) N2 and 16 of 821 (1.94%) N3 disease (TNM 8). Forty-eight (5.85%) cases had missing staging data. A total of 88 of 821 (10.7%) had equivocal iENE. Of the 821

Table 1 Characteristics of patients with HPV+ by iENE status

Variable		Overall n (%)	iENE n (%)		*P* value*
			Negative (n = 601)	Positive (n = 220)	
Biological sex (n = 821)	Male	651 (78.53)	459 (76.37)	185 (84.09)	**.017**
	Female	178 (21.47)	142 (23.63)	35 (15.91)	
Smoking status (n = 821)	Never	260 (31.67)	203 (33.78)	57 (25.91)	**.014**
	Previous	328 (39.95)	237 (39.43)	91 (41.36)	
	Current	174 (21.19)	127 (21.13)	47 (21.36)	
	Unknown	59 (7.19)	34 (5.66)	25 (11.36)	
Recruiting center (n = 821)	Australia — Royal Adelaide Hospital	85 (10.35)	48 (7.99)	37 (16.82)	**<.001**
	Denmark — University of Copenhagen	125 (15.23)	123 (20.47)	2 (0.91)	
	France	14 (1.71)	8 (1.33)	6 (2.73)	
	Germany — Cologne	76 (9.26)	61 (10.15)	15 (6.82)	
	Germany — Munich	42 (5.12)	22 (3.66)	20 (9.09)	
	Netherlands — Amsterdam UMC	223 (27.16)	162 (26.96)	61 (27.73)	
	Netherlands — Amsterdam NKI/UVA	20 (2.44)	5 (0.83)	15 (6.82)	
	Spain — Institut Catala dOncologia	11 (1.34)	10 (1.66)	1 (0.45)	
	Switzerland — Zurich	41 (4.99)	34 (5.66)	7 (3.18)	
	UK — University Hospitals Birmingham	33 (4.02)	27 (4.49)	6 (2.73)	
	USA — Icahn School of Medicine at Mount Sinai	33 (4.02)	23 (3.83)	10 (4.55)	
	USA — Methodist Eastbrook Cancer Centre	41 (4.99)	15 (2.50)	26 (11.82)	
	USA — University of Pittsburgh Medical	77 (9.38)	63 (10.48)	14 (6.36)	
Performance status grade/score (ECOG -Karnofsky) (n = 442)	ECOG score 0 = Karnofsky PS score 90-100	277 (62.67)	234 (66.1)	43 (48.86)	**.004**
	ECOG score 1 = Karnofsky PS score 70-80	123 (27.83)	94 (26.55)	29 (32.95)	
	ECOG score 2 = Karnofsky PS score 50-60	31 (7.01)	20 (5.65)	11 (12.5)	
	ECOG score 3 = Karnofsky PS score 30-40	11 (2.49)	6 (1.69)	5 (5.68)	
Oropharynx subsite (n = 640)	Base of tongue	234 (36.56)	164 (34.38)	70 (42.94)	.253
	Posterior pharyngeal wall	9 (1.41)	7 (1.47)	2 (1.23)	
	Tonsil	379 (59.22)	293 (61.43)	86 (52.76)	
	Other oropharynx	18 (2.81)	13 (2.73)	5 (3.07)	

(*Continued*)

Table 1 (*Continued*)

Variable		Overall n (%)	iENE n (%)		*P* value*
			Negative (n = 601)	Positive (n = 220)	
TNM8 tumor category (n = 774)	T0	9 (1.16)	7 (1.24)	2 (0.95)	**.014**
	T1	244 (31.52)	195 (34.64)	49 (23.22)	
	T2	295 (38.11)	200 (35.52)	95 (45.02)	
	T3	71 (9.17)	50 (8.88)	21 (9.95)	
	T4	83 (10.72)	54 (9.59)	29 (13.74)	
	TX	72 (9.30)	57 (10.12)	15 (7.11)	
TNM8 nodal category (n = 773)	N0	169 (21.86)	156 (27.71)	13 (6.19)	**<.001**
	N1	501 (64.81)	351 (62.34)	150 (71.43)	
	N2	87 (11.25)	51 (9.06)	36 (17.14)	
	N3	16 (2.07)	5 (0.89)	11 (5.24)	
Treatment (n = 821)	Surgery Alone	203 (24.73)	175 (29.12)	28 (12.73)	**<.001**
	Surgery + RT	149 (18.15)	119 (19.80)	30 (13.64)	
	Surgery + CRT	176 (21.44)	113 (18.80)	63 (28.64)	
	RT Alone	71 (8.65)	63 (10.48)	8 (3.64)	
	CRT	221 (26.92)	130 (21.63)	91 (41.36)	
	Other	1 (0.12)	1 (0.17)	0 (0.00)	
Age at diagnosis (y) (n = 821)	Mean (SD)	59.86 (9.40)	59.91(9.70)	59.73 (8.52)	.808
Time between scan and treatment (n = 279)	≤4 wks	153 (54.84)	109 (58.92)	44 (46.81)	.097
	>4 and ≤8 wks	101 (36.20)	63 (34.05)	38 (40.43)	
	>8 wks	25 (8.96)	13 (7.03)	12 (12.77)	
pENE (n = 510)	Negative	318 (62.35)	285 (72.34)	33 (28.45)	**<.001**
	Positive	192 (37.65)	109 (27.66)	83 (71.55)	
Follow-up time (mo) (n = 819)	Median (IQR)	50.5 (28.7, 82.0)	51.3 (29.6, 83.0)	48.0 (26.0, 76.5)	.251
Survival outcome (n = 821)	Dead	128 (15.59)	84 (13.98)	44 (20.00)	.093
	Alive	693 (84.41)	517 (86.02)	176 (80.00)	

Abbreviations: CRT = chemoradiation therapy; iENE = imaging-detected extranodal extension; PS = performance status; RT = radiation therapy.
* The *P* value is calculated by *t* test or Wilcoxon rank-sum test for numerical variables and χ^2 test for categorical variables.

patients, 203 (24.7%) had surgery alone; 71 (8.70%) RT alone; 221 (26.9%) CRT; 149 (18.2%) surgery and RT; and 176 (21.4%) surgery and CRT. Median follow-up was 4.18 years (IQR, 2.4-6.8 years).

A total of 220 of 821 (26.8%) had iENE+ disease. Of those who had surgery and histopathological data, 192 of 510 (37.6%) had pENE+ nodal disease. In comparison with cases that were iENE negative, iENE+ cases were of the same mean age (59.7 vs 59.9 years), but had a higher proportion of men (84.1% vs 76.5%, $P = .019$) (Table 1); fewer never smokers (25.9% vs 33.8%, $P = .014$); more patients with poor performance status 2 and 3 (18.2% vs 6.37%, $P = .004$); higher proportion of T3 and T4 disease (23.7% vs 18.47%, $P = .004$); and a higher proportion of N+ disease (93.81% vs 72.68%, $P < .001$). Significantly more patients with iENE received CRT (41.4% vs 21.6%) or surgery plus CRT (28.6 vs 18.8%), and fewer patients with iENE received unimodal (surgery or RT alone) treatment (16.3% vs 39.6%) or surgery plus RT alone (13.6 vs 19.8%) than those without iENE ($P < .001$). The likelihood of receiving triple therapy when having surgery was 1.87 if iENE was present compared with if it was absent, with 52.2% of patients with iENE who received surgery also having postoperative CRT (indicating pENE) compared with 27.8% of those without iENE ($P < .001$).

Diagnostic accuracy

Five hundred twenty-seven patients received surgery, 6 were missing iENE and 21 missing pENE, resulting in a total of 638 neck sides available for the diagnostic test accuracy analysis (Fig. E1). 56.8% (109 of 192) of patients with pENE (pENE+) were misclassified as negative on iENE (iENE−), and 109 of 394 (27.7%) of patients who were iENE negative had pENE on histopathology. Therefore, sensitivity and specificity of iENE in identifying any type of pENE was 44.5% (95% CI, 37.8%-51.4%) and 87.6% (95% CI, 84.1%-90.6%) respectively (Table E1). Positive and negative likelihood ratios were 3.6 (95% CI, 2.7-4.8) and 0.6 (95% CI, 0.6-0.7) respectively. The positive predictive value was 65.1% (95% CI, 58.2%-71.5%), the negative predictive value was 75.3% (95% CI, 72.3%-77.5%), and the area under the curve (AUC) was 0.417 (95% CI, 0.370-0.465).

The results of sensitivity analyses by diagnostic criteria (including unequivocal cases as positive), by pENE > 2 mm and by center, are provided in detail in File E1, Table E2 (A, B), and Table E3. There were statistically significant differences in sensitivity and specificity between several centers (Tables E4 and E5), albeit some centers only had a small number of cases. Interactions in results between imaging modality and radiologist specialization or between modality and treatment center could not be reliably undertaken because of sample sizes.

When subgroup analysis was undertaken by modality (Fig. 1 and Table E6), the sensitivity in the 341 patients who had CT scans was significantly higher (47.45%; 95% CI, 38.86%-56.15%) than that in 196 patients with MRI (18.60%; 95% CI, 8.39-33.40; $P < .001$), but the specificity was significantly lower (78.43%; 95% CI, 72.15%-83.87%) versus 96.73% (95% CI, 92.54%-98.93%; $P < .001$). In the 64 patients who had both CT and MRI and were positive on at least 1 of them, both sensitivity and specificity increased significantly: 84.6% (95% CI, 65.1%-95.6%; $P < .001$) and 94.5% (95% CI, 82.3%-99.4%; $P = .022$) compared with CT alone respectively, and sensitivity was significantly higher (18.6%; 95% CI, 8.4%-33.4%; $P < .001$) compared with MRI alone. There was no difference in specificity between CT and MRI together and MRI alone ($P = .628$). The AUC for combined CT and MRI (0.824; 95% CI, 0.724-0.924) was much higher than CT (0.423; 95% CI, 0.361-0.486) or MRI alone (0.183; 95% CI, 0.102-0.264). There was a statistically significant interaction between iENE and modality in the prediction of pENE ($P = .002$, Table E6).

When subgroup analysis was undertaken by expertise (Fig. 1), head and neck specialists showed significantly higher specificity (89.14%; 95% CI, 85.69%-91.99% vs 46.7%; 95% CI, 21.3%-73.4%; $P < .001$) and there was a statistically significant interaction between iENE and expertise in the prediction of pENE ($P = .028$, Table E7), but there were no significant differences in sensitivity, positive and negative predictive values, or AUC between specialists and nonspecialists. There were also no statistically significant differences in any of the parameters or in the interaction term when subgroup analysis was

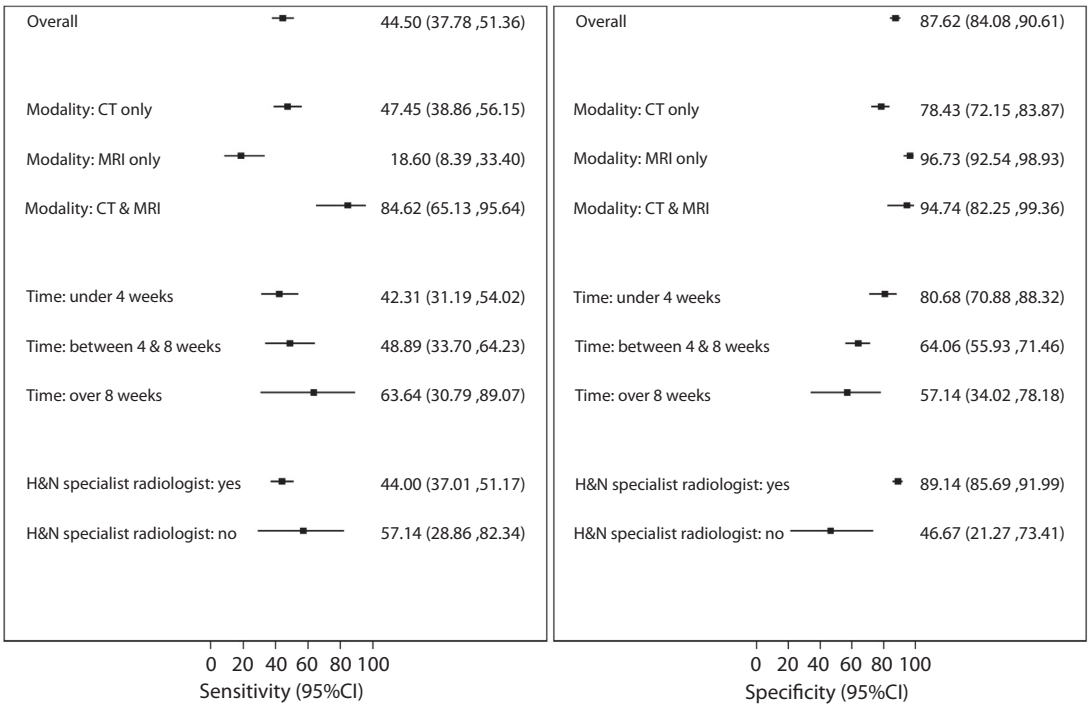

Fig. 1. Forrest plot of sensitivity (left) and specificity (right), overall and across diagnostic modalities, interval between diagnosis and treatment and the involvement of a radiologist.

undertaken by scan to treatment interval (4 weeks, 4-8 weeks, and >8 weeks) (Fig. 1, Tables E8 and E9).

Survival

The median 5-year OS of the cohort was 90.4% (SE, 1.1%) and the 5-year disease-free survival (DFS) was 83.9% (SE, 1.34%) (Fig. 2). After removing 167 patients with missing follow-up time, smoking status, TNM eighth edition stage or tumor stage, overall and DFS analysis was conducted on 654 patients. Cases with iENE exhibited significantly worse unadjusted 5-year OS compared with iENE negative 79.4% (95% CI, 71.9-85.1) vs 88.1% (95% CI, 84.8-90.6; $P = .025$) (Table E10). iENE-positive cases also demonstrated worse 5-year DFS (74.9%; 95% CI, 67.6-80.8) vs (81.6; 95% CI, 78.0-84.8; $P = .034$). Differences in locoregional control (88.2; 95% CI, 82.5-92.1) vs (92.0; 95% CI, 89.2-94.1; $P = .129$) and distant metastasis-free survival (92.6; 95% CI, 87.5-95.7) vs (94.4; 95% CI, 91.8-96.1; $P = .193$) were not statistically significant.

On multivariate analysis, iENE positivity was *not* a statistically significant predictor of OS, with an adjusted hazards ratio (aHR) for OS of 1.50 (95% CI, 0.97-2.32; $P = .071$) compared with iENE-negative cases (Table 2). The following factors were identified as statistically significant predictors of OS: age (aHR, 1.05; 95% CI, 1.03-1.08; $P < .001$), ever smoked status (aHR, 1.63; 95% CI, 1.03-2.58; $P = .036$), T3/T4 T-category (aHR, 2.22; 95% CI, 1.45-3.39; $P < .001$), and N3 nodal category (aHR, 5.14; 95% CI, 2.02-13.11; $P = .001$) (Table 2).

For the sake of comparison, in multivariate analysis pENE positivity was a statistically significant predictor of OS, with aHR 3.11 (95% CI, 1.52-6.38; $P = .002$, n = 360). Age, T3/T4 T-category, and N3 nodal category remained as statistically significant predictors of OS.

Sensitivity analysis of equivocal iENE

We did a *post hoc* analysis of the multivariate analysis models with equivocal iENE counted as iENE positive (iENE eqivocal-positive), because the sensitivity analyses had suggested that this significantly increased sensitivity and to a lesser degree decreased specificity. When incorporated into the multivariate model, iENE eqivocal-positive was a significant predictor of OS (HR, 1.60; 95% CI, 1.04- 2.46; $P = .032$). The DFS model results and sensitivity analyses are presented in the supplementary results section and Table E11.

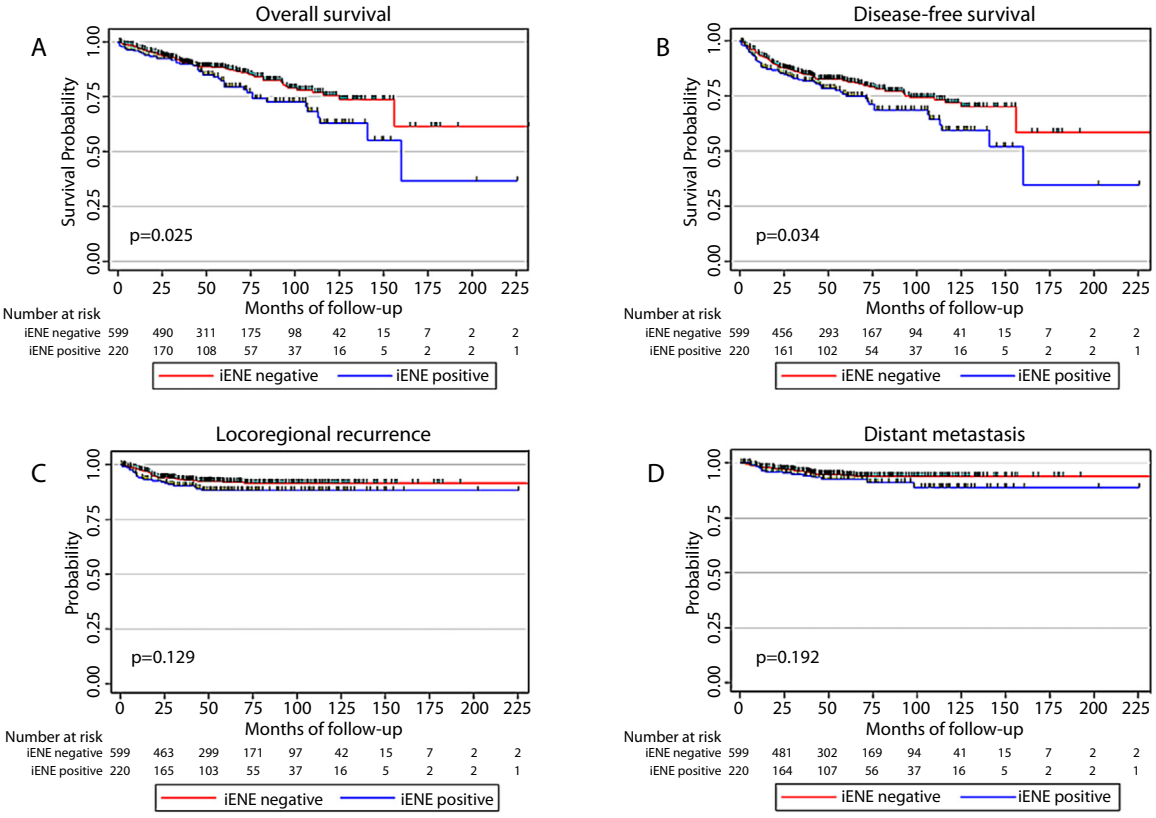

Fig. 2. Outcomes by iENE status. (A) Overall survival. (B) Disease-free survival. (C) locoregional recurrence. (D) distant metastasis. Shown are the *P* values for a difference in 5-year survival probability (for details see supplementary Table E10).

Table 2 Multivariable analysis of overall survival in the population of cases with complete data on all included variables (n = 654)

		Harrell's C-index: 0.718	
Variable	**Hazard ratio (95% CI)**	***P* value**	**Global *P* value**
Age	1.05 (1.03-1.08)		**<.001**
Sex (female ref.)			.072
Male	1.63 (0.96-2.76)		
Smoking status (never ref.)			**.036**
Ever smoked	1.63 (1.03-2.58)		
Nodal category (N0 ref.)			**<.001**
N1	0.89 (0.52-1.53)	.668	
N2	1.10 (0.53-2.26)	.805	
N3	5.14 (2.02-13.11)	**.001**	
Tumor category (T0-T2 ref.)			**<.001**
T3-T4	2.22 (1.45-3.39)		
Treatment (Surgery alone ref.)			.722
Surgery + RT	0.70 (0.34-1.42)	.320	
Surgery + CRT	0.69 (0.32-1.45)	.327	
RT alone	0.90 (0.44-1.83)	.769	
CT alone	0.65 (0.33-1.26)	.202	
Other	-	-	
iENE (Negative ref.)			.071
iENE positive	1.50 (0.97-2.32)		

Abbreviations: CRT = chemoradiation therapy; Ref = reference; RT = radiation therapy.

Proposed amendments to TNM eighth classification

Finally, we assessed the 2 proposals for amendments to the TNM eighth classification (Fig. 3). Neither new proposal produced large improvements in nodal category or overall stage to the current TNM 8 system. The results for nodal categories and overall stage are detailed in File E1, Fig. 1, and Tables E12 and E13.

Discussion

Our study demonstrates that in a multinational real-world setting, iENE exhibits modest negative predictive value with significant variability in sensitivity and specificity across centers, despite indications of its influence on treatment selection in routine practice. iENE was not a statistically significant independent prognostic factor for OS or DFS in both the overall cohort and surgically treated subgroup. We also evaluated 2 new proposals to integrate iENE into the TNM 8 classification system, finding neither substantially improved prognostication in real-world practice.

Notably, we identified that the accuracy and prognostic utility of iENE improved by several interventions. Sensitivity and specificity increased when both CT and MRI were employed, compared to either modality alone. Specialist head and neck cancer radiologists achieved higher specificity than nonspecialists. Additionally, using inclusive diagnostic criteria (considering both positive and equivocal cases as iENE+) improved sensitivity and prognostic power but reduced specificity. The interval between imaging and treatment did not significantly impact accuracy, possibly because of iENE's limited prognostic relevance in HPV + OPC, the mitigating effect of adjuvant therapy, or a type II statistical error from small sample sizes.

Sensitivity and specificity varied widely across centers, reflecting inconsistencies also noted in previous meta-analyses.[6,7] This variability likely stems from the absence of consensus on optimal imaging modalities and diagnostic criteria for iENE. Most centers used CT, which showed higher sensitivity but lower specificity than MRI, consistent with a recent meta-analysis.[6] We identified that combining

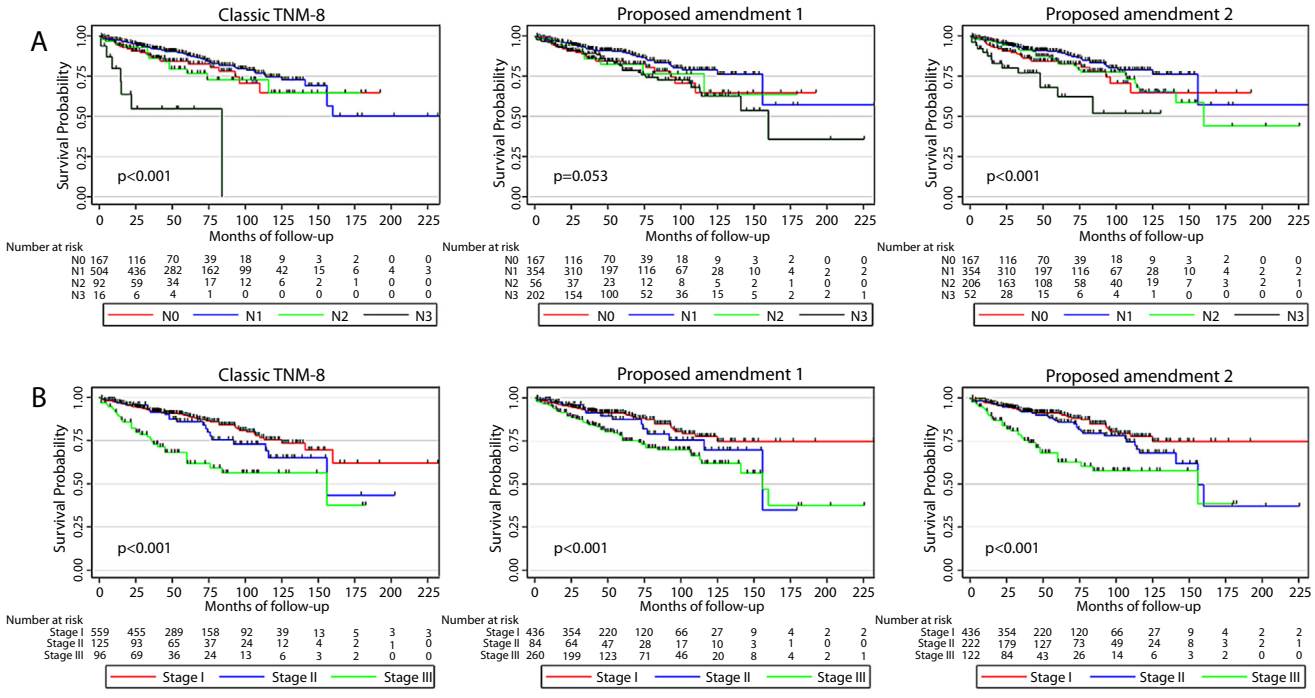

Fig. 3. Nodal (A) and overall staging (B) of patients (numbers [%]) according to (from left to right) the classic TNM-8 and 2 new proposals. Shown are the *P* values for a difference in 5-year survival probability (for details see supplementary Table E12).

CT and MRI significantly enhanced diagnostic accuracy, warranting further validation.

The lack of standardized diagnostic criteria for iENE contributes to variability. Centers used their routine clinical criteria. Sensitivity analyses revealed that inclusive criteria improved sensitivity and prognostic relevance, potentially highlighting radiologists' cautious approach because of its effect on treatment implications. Consensus diagnostic criteria could improve accuracy, reporting confidence, and have been shown to reduce interobserver variability,[10] enhancing the prognostic value and clinical utility of iENE. The Head Neck Cancer International Group (HNCIG; www.hncig.org) recently published consensus criteria and a 4-tier classification for iENE,[11] ranging from irregular margins and/or perinodal fat invasion (grade 1), matted or conglomerate nodes (grade 2) and invasion in to adjacent structures, eg, skin or muscle (grade 3). Emerging radiomics and machine learning may further standardize iENE assessment.[12] HNCIG has embarked on the Artificial Intelligence in Head and Neck Cancer initiative to collect data sets that will help enable the development and validation of artificial intelligence-driven algorithms for iENE identification among others.

The surgical subset is inherently biased because they are less likely to have iENE. Sensitivity and specificity may be higher and more consistent in patients with overt iENE who are treated with nonsurgical approaches. Nevertheless, the "absence of iENE" is increasingly being used in trials and in daily practice around the world to select patients for surgical treatment, as opposed to CRT.[2-4] In the surgical cohort, iENE showed limited iENE accuracy compared with pENE. iENE-negative status does not reliably predict absence of pENE, with 28% of iENE-negative patients having pENE on histopathology. This discrepancy contributes to the high incidence (~30%) of triple therapy in transoral surgery patients,[2,13] despite initial plans for single-modality treatment. Conversely, 28% of iENE-positive cases lacked pENE, potentially denying them less intensive treatment. iENE's prognostic power was modest (HR, 1.5) with a ~9% OS difference between iENE+ and iENE- groups, contrasting with stronger associations in smaller, single-center studies with single reporting radiologists.[5,9,14,15] This may result from study publication bias of positive studies, study design differences, statistical limitations, or the differences in the treatment of ENE, resulting in differences in the mitigating effect of adjuvant treatments.

Study limitations include its retrospective design, incomplete performance status data, and potential confounding from incorporating scans over 20 years old, which may have had poorer resolution. Because of limitations of data and privacy, we were unable to analyze the effect of date of collection on accuracy. HPV + OPC was defined by p16 positivity, although some p16+ cases are HPV-negative with poorer outcomes, possibly inflating iENE's prognostic value.[8]

Radiologists used varied criteria without standardized training. We believe that this is in fact a strength of our study, reflecting real-world variability and lower sensitivity compared with systematic reviews of single-center studies. This variability underscores the need for standardized consensus criteria, such as that published by HNCIG and endorsed by 22 organizations representing 39 countries.[11,16]

Our study's other strengths include its large, global, diverse cohort, enabling robust subgroup analyses and real-world applicability.

Given iENE is already being used for selecting patients for surgical treatment in clinical practice, international collaboration is needed to develop and validate consensus diagnostic protocols, as well as develop and validate artificial intelligence models. Until then, clinicians should exercise caution when using iENE for treatment selection.

References

1. Cracchiolo JR, Baxi SS, Morris LG, et al. Increase in primary surgical treatment of T1 and T2 oropharyngeal squamous cell carcinoma and rates of adverse pathologic features: National Cancer Data Base. *Cancer* 2016;122:1523-1532.
2. Ferris RL, Flamand Y, Weinstein GS, et al. Phase II randomized trial of transoral surgery and low-dose intensity modulated radiation therapy in resectable p16+ locally advanced oropharynx cancer: an ECOG-ACRIN cancer research group trial (E3311). *J Clin Oncol* 2022;40:138-149.
3. Hargreaves S, Beasley M, Hurt C, Jones TM, Evans M. Deintensification of adjuvant treatment after transoral surgery in patients with human papillomavirus-positive oropharyngeal cancer: the conception of the PATHOS study and its development. *Front Oncol* 2019;9:936.
4. Homer JJ, Winter SC. Head and Neck Cancer: United Kingdom National Multidisciplinary Guidelines, sixth edition. *J Laryngol Otol* 2024;138(suppl 1) S1-s224.
5. Benchetrit L, Torabi SJ, Givi B, Haughey B, Judson BL. Prognostic significance of extranodal extension in HPV-mediated oropharyngeal carcinoma: a systematic review and meta-analysis. *Otolaryngol Head Neck Surg* 2021;164:720-732.
6. Morey T, Stern C, Foreman A, Hodge JC. Accuracy of imaging modalities at detecting extracapsular spread of cervical lymph node metastases in HPV-associated oropharyngeal cancer: a systematic review protocol of diagnostic test accuracy. *JBI Evid Synth* 2022;20:189-195.
7. Park SI, Guenette JP, Suh CH, et al. The diagnostic performance of CT and MRI for detecting extranodal extension in patients with head and neck squamous cell carcinoma: a systematic review and diagnostic meta-analysis. *Eur Radiol* 2021;31:2048-2061.
8. Mehanna H, Taberna M, von Buchwald C, et al. Prognostic implications of p16 and HPV discordance in oropharyngeal cancer (HNCIG-EPIC-OPC): a multicentre, multinational, individual patient data analysis. *Lancet Oncol* 2023;24:239-251.
9. Huang SH, O'Sullivan B, Su J, et al. Prognostic importance of radiologic extranodal extension in HPV-positive oropharyngeal carcinoma and its potential role in refining TNM-8 cN-classification. *Radiother Oncol* 2020;144:13-22.
10. Hoebers F, Yu E, O'Sullivan B, et al. Augmenting inter-rater concordance of radiologic extranodal extension in HPV-positive oropharyngeal carcinoma: a multicenter study. *Head Neck* 2022;44:2361-2369.
11. Henson C, Abou-Foul AK, Yu E, et al. Criteria for the diagnosis of extranodal extension detected on radiological imaging in head and neck cancer: HNCIG international consensus recommendations. *Lancet Oncol* 2024;25:e297-e307.
12. Kann BH, Likitlersuang J, Bontempi D, et al. Screening for extranodal extension in HPV-associated oropharyngeal carcinoma: evaluation of a CT-based deep learning algorithm in patient data from a multicentre, randomised de-escalation trial. *Lancet Digit Health* 2023;5:e360-e369.
13. Haughey BH, Sinha P, Kallogjeri D, et al. Pathology-based staging for HPV-positive squamous carcinoma of the oropharynx. *Oral Oncol* 2016;62:11-19.
14. Billfalk-Kelly A, Yu E, Su J, et al. Radiologic extranodal extension portends worse outcome in cN+ TNM-8 stage I human papillomavirus −mediated oropharyngeal cancer. *Int J Radiat Oncol Biol Phys* 2019;104:1017-1027.
15. Thompson LDR, Burchette R, Iganej S, Bhattasali O. Oropharyngeal squamous cell carcinoma in 390 patients: analysis of clinical and histological criteria which significantly impact outcome. *Head Neck Pathol* 2020;14:666-688.
16. Abou-Foul AK, Henson C, Chernock RD, et al. Standardised definitions and diagnostic criteria for extranodal extension detected on histopathological examination in head and neck cancer: Head and Neck Cancer International Group consensus recommendations. *Lancet Oncol* 2024;25:e286-e296.