



## Determining Disease Extent

# Radiologic-Pathologic Correlation of Extranodal Extension in Patients With Squamous Cell Carcinoma of the Oral Cavity: Implications for Future Editions of the TNM Classification



Abdullah Almulla, MD,\* Christopher W. Noel, MD,<sup>†</sup> Lin Lu, MSc,<sup>‡</sup>  
Wei Xu, PhD,<sup>‡</sup> Brian O'Sullivan, MB BCh,<sup>†,§</sup> David P. Goldstein, MD,<sup>†</sup>  
Andrew Hope, MD,<sup>§</sup> Bayardo Perez-Ordóñez, MD,<sup>||</sup> Ilan Weinreb, MD,<sup>||</sup>  
Jonathan Irish, MD, MSc,<sup>†</sup> Patrick Gullane, MB,<sup>\*,†</sup>  
Douglas Chepeha, MD,<sup>†</sup> Li Tong, MD, BSc,<sup>§</sup> Eugene Yu, MD,\*  
and Shao Hui Huang, MD, MSc<sup>§</sup>

*A cura di Carlotta Becherini*

## **Abstract**

**Purpose:** Valutare l'accuratezza ed il valore prognostico della presenza di estensione extranodale radiologica (rENE) versus l'estensione extranodale patologica (pENE) nei pazienti affetti da tumore a cellule squamose del cavo orale (OSCC).

**Methods and Materials:** Una review retrospettiva è stata condotta su tutti i pazienti con nuova diagnosi di OSCC, sottoposti a dissezione linfonodale cervicale, afferiti nel nostro istituto dal 2010 al 2015, di cui fossero disponibili TC o MRI preoperatorie. Due neuroradiologi dedicati allo studio della regione del Testa-Collo hanno revisionato, in maniera indipendente, le immagini per la presenza di rENE (descritta come margini linfonodali mal-definiti) e senza essere a conoscenza del precedente referto diagnostico. È stato valutato l'impatto del tempo tra l'imaging e la chirurgia, della modalità di imaging, della concordanza del rENE intra-classificatore ed inter-classificatore. È stata valutata l'accuratezza diagnostica del rENE versus pENE. È stata comparata la sopravvivenza globale (OS) tra i pazienti con e senza rENE. Il valore prognostico del rENE è stato valutato tramite analisi multivariata.

**Results:** Tra i 508 pazienti inclusi, sono stati identificati rENE e pENE in 57 e 121 casi, rispettivamente. L'accuratezza diagnostica fra rENE versus pENE è stata identica (73%) nei casi in cui l'intervallo fra l'imaging e la chirurgia era di 4 settimane (n=276) e tra 4 ed 8 settimane (n=207), ma diminuiva (48%) per intervalli superiori ad 8 settimane (n=25). La TC ha dimostrato una più alta accuratezza nella valutazione del rENE rispetto alla RM (80% vs 63%, P= .011). La concordanza inter-operatore ed intra-operatore (n =93) è stata, rispettivamente, buona (k = 0.79) ed eccellente (k = 0.94). Escludendo i 25 casi con un intervallo imaging-chirurgia >8 settimane, nei restanti 483 casi



la sensibilità, la specificità, il valore predittivo positivo e negativo del rENE versus pENE erano rispettivamente 52%, 96%, 93, e 66%. I pazienti che presentano rENE (n = 55) hanno una OS inferiore rispetto a quelli senza rENE (n = 202), ed entrambi hanno un OS minore ai pazienti con linfonodi negativi (n = 226) (OS a 3 anni: 31% vs 68% vs 81%,  $P < .001$ ). L'analisi multivariata, aggiustata per età, categoria del T, categoria dell'N e performance status ha confermato il valore prognostico del rENE rispetto all'OS (hazard ratio 3.3, 95% confidence interval 2.4- 5.3,  $P < .001$ ).

**Conclusions:** Questo ampio studio di coorte evidenzia un'alta specificità ma una bassa sensibilità del rENE per il pENE. Analogamente al pENE, la presenza di rENE è associata con una sopravvivenza ridotta nei OSCC.

L'estensione extranodale (ENE) è definita come la diffusione delle cellule tumorali oltre la capsula di un linfonodo metastatico (LN) verso i tessuti perinodali. Nei pazienti con i tumori della regione testa-collo (HNC), l'ENE è uno dei fattori prognostici più importanti circa l'aggressività della malattia, coinvolgendo gli esiti del trattamento e la sopravvivenza. Riconoscendo il suo significato prognostico, l'ottava edizione del manuale AJCC per la classificazione TNM, adesso include questo parametro per la classificazione dell'N negli HNC non correlati a virus, inclusi gli OSCC. Sebbene il gold standard per la diagnosi di ENE sia basato su evidenze patologiche, la presenza radiologica di "margini nodali mal-definiti" è considerato un classico reperto caratteristico di ENE. Tuttavia, rimane ancora poco chiaro se ENE possa essere identificato in maniera affidabile nell'imaging preoperatorio e quale importanza prognostica abbia, dal momento che gli studi su questo argomento mostrano risultati inconcludenti. Le linee guida attuali si basano sulla conferma patologica della presenza di ENE (pENE), piuttosto che quella radiologica (rENE), per lo staging e la scelta diagnostica. A causa di una scarsità di dati robusti a conferma dell'affidabilità del rENE rispetto alla valutazione del pENE, l'ottava edizione del manuale di stadiazione AJCC utilizza criteri più stringenti per la definizione clinica dell'ENE (cENE). Solo coloro che abbiano una evidenza netta di presenza macroscopica di ENE (es. coinvolgimento dermico od invasione dei tessuti molli con fissità sui muscoli sottostanti o le strutture adiacenti oppure segni clinici di interessamento nervoso) sono considerati rappresentare il cENE. E' possibile che la classificazione N possa essere ulteriormente migliorata se la presenza di ENE potesse essere identificata in maniera affidabile ed accurata sulla TC o sulla MRI (rispetto alla pENE) da differenti operatori (inter-operatore) e/o dallo stesso operatore a diversi time points (intra-operatore) e se questo recasse con sé una significatività prognostica. Pertanto, è stato condotto questo studio su una coorte di pazienti affetti da OSCC sottoposti a dissezione linfonodale presso il nostro istituto. Questo studio è stato progettato per affrontare i seguenti problemi: il grado di correlazione tra l'evidenza radiologica e patologica di ENE, il valore prognostico della rENE per l'OS, il valore dell'aggiunta delle caratteristiche radiologiche sul possibile miglioramento nella previsione del riscontro di pENE ed infine la concordanza del rENE inter ed intra operatore.

## **Materiali e Metodi**

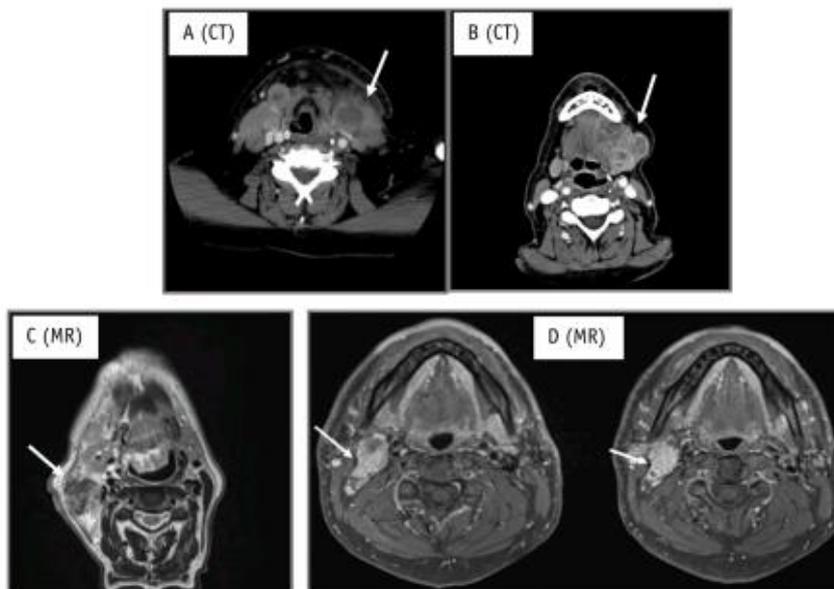
### *Popolazione dello studio*

Si tratta di uno studio retrospettivo di revisione dell'imaging TC o MRI preoperatorio di tutti i pazienti con nuova diagnosi di OSCC, sottoposti a dissezione linfonodale cervicale, presso l'istituto di riferimento dal 2010 al 2015. I criteri di esclusione erano: scarsa qualità dell'imaging, un ritardo nel tempo alla chirurgia (>12 settimane tra l'imaging preoperatorio e la chirurgia), o l'assenza di imaging preoperatorio. Sono state utilizzate TC con mezzo di contrasto seguendo un protocollo standardizzato (slice di 2mm di spessore). Sono state effettuate ricostruzioni sagittali e coronali. Tutti i pazienti hanno ricevuto mdc iv. Le immagini MR sono state acquisite da macchinari MRI da 1.5T o da 3T, utilizzando un protocollo standard per la regione testa-collo che include sequenze T1,

T2-pesate, T2 con soppressione del tessuto grasso e T1 post-gadolino con soppressione del tessuto grasso, mentre non include immagini con il coefficiente apparente di diffusione o pesate sulla diffusione. Entrambe le immagini TC e MRI sono state ottenute a partire da un livello subito al di sopra della sommità dei ventricoli laterali fino ad arrivare inferiormente all'apice della cavità toracica. Le immagini radiologiche sono state revisionate utilizzando un sistema disponibile in commercio (Coral PACS, Canada).

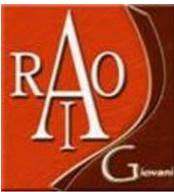
### *Informazioni Cliniche e Radiologiche*

I dati clinici, inclusi i rilievi patologici, sono stati ottenuti dal database prospettico istituzionale Head-and Neck-Anthology-of-Outcomes. L'evidenza radiologica di linfonodi conclamati (rN+) e di rENE è stata revisionata da A.A., il quale era all'oscuro circa i referti patologici. Sulla base del calcolo della potenza statistica, sono stati considerati sufficienti 80 casi al fine di garantire almeno l'85% di potenza statistica per rilevare differenze statisticamente significative nella concordanza intra ed inter-operatore. Tuttavia, al fine di garantire un numero sufficiente di pazienti eleggibili, sono stati selezionati in maniera casuale 100 casi per la valutazione dell'affidabilità della stima di rN+, rENE, ed aggregati linfonodali. rENE è stato definito come la presenza di margini linfonodali mal-definiti. Gli aggregati linfonodali sono stati definiti come la presenza di 3 linfonodi adiacenti, contigui l'uno con l'altro, con la perdita di piano di clivaggio. La Figura 1 riproduce la tipica presentazione all'imaging TC ed MRI di rENE. La presenza di linfonodi coinvolti patologicamente (pN+) e di pENE è stata registrata sulla base dei repost patologici effettuati al momento del trattamento iniziale, ed è stata effettuata da 2 patologi esperti di testa-collo (B.P.O. e I.W.). La presenza di pENE è stata definita durante la valutazione patologica come l'infiltrazione delle cellule tumorali oltre la capsula dei linfonodi metastatici. Il grado di pENE (es, l'estensione di pENE oltre la capsula linfonodale) non era disponibile nel momento in cui è stato effettuato lo studio.



**Fig. 1.** Typical radiologic appearance of extranodal extension on computed tomography (CT) and magnetic resonance (MR) images. Contrast-enhanced CT shows the presence of rENE (white arrows) in necrotic left level II nodes with marked peripheral ill definition (A), and matted lymph nodes in left 1B region (B); T2-weighted MR image shows rENE (white arrows) in a right 2A node with ill-defined margins (C) and a matted node in right level IIA (shows in 2 continuous planes) (D).

### *Analisi Statistica*



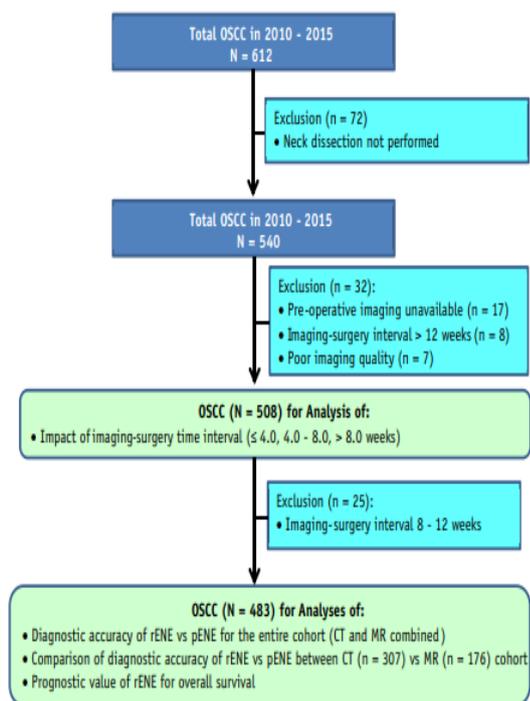
E' stata effettuata un'analisi esplorativa al fine di identificare un accettabile intervallo di tempo tra imaging preoperatorio e chirurgica attraverso la comparazione dell'accuratezza diagnostica degli rN+ versus pN+ e della rENE versus pENE stratificando per intervalli di 4, 4-8, e 8-12 settimane. Escludendo i pazienti con intervalli di tempo non accettabili, i restanti pazienti sono stati ritenuti elegibili per le analisi successive. Le analisi conseguenti esplorano l'accuratezza diagnostica del rENE versus pENE, la comparazione della TC versus MRI per il rENE, il valore prognostico del rENE e la concordanza inter ed intra operatore per il rENE. La sensibilità, la specificità, il valore predittivo positivo (PPV), il valore predittivo negativo (NPV) e l'accuratezza sono state calcolate per rNp versus pNp e per rENE versus pENE sia nell'intera coorte, che in due coorti separate per la TC e la MRI. Z-score test è stato usato per il confronto tra le accuratezze diagnostiche. Cohen's kappa test è stato utilizzato per esaminare la concordanza intra ed inter-operatore nella valutazione di rNp, rENE, e di aggregati linfonodali. La regressione logistica è stata effettuata al fine di calcolare l'odds ratio (OR) del rENE per il pENE, con la presenza di pENE come "evento" e la sua assenza come valore di riferimento. L'analisi di Kaplan-Meier è stata utilizzata per calcolare l'OS, in essa l'"evento" è stato considerato la morte da ogni causa. Il Time-to-event è stato calcolato dalla data della chirurgia. Tre modelli di analisi multivariata (MVA) sono stati condotti per valutare il valore prognostico di rENE, pENE, and both ENE e pENE, adattati per categoria T, categoria N, Zubrod PS ed età. Tutti i test erano 2-sided, ed i risultati sono stati considerati significativi per valori di P .05.

## **Risultati**

*L'impatto dell'intervallo tra l'imaging preoperatorio e la chirurgia.*

Dei 612 pazienti affetti da OSCC consecutivi, 540 (88%) sono stati sottoposti a dissezione linfonodale cervicale. Escludendo 32 casi per mancanza di imaging pre-operatorio (n = 17),

intervallo di tempo prolungato (>12 settimane) tra imaging e chirurgia (n = 8), e scarsa qualità dell'imaging (n = 7), i restanti 508 casi sono stati ritenuti elegibili e sottoposti all'analisi esplorativa circa l'efficacia diagnostica del rENE versus pENE attraverso differenti intervalli di tempo di imaging-chirurgia. I 508 casi sono stati stratificati in 3 sottogruppi: 4 settimane (n = 276), da 4 a 8 settimane (n = 208), e da 8 a 12 settimane (n = 25). L'accuratezza diagnostica del rENE versus pENE era identica (73%) per i sottogruppi con intervallo di 4 settimane e 4-8 settimane ma inferiore per il sottogruppo 8-12 settimane (48%). Seppur non statisticamente significativo (probabilmente legato alla piccola dimensione del campione) se confrontato con il sottogruppo di 8 settimane (P = .457), il sottogruppo 8-12 settimane (n=25) è stato escluso ed i rimanenti 483 casi sono stati valutati nelle analisi successive (Fig. 2).



**Fig. 2.** Derivation of study population and analyses schema. Abbreviations: CT = computed tomography; MR = magnetic resonance; OSCC = oral cavity squamous cell carcinoma; pENE = pathologic extranodal extension; rENE = radiologic extranodal extension.

### Concordanza di rENE intra ed inter-operatore.

Dei 100 casi selezionati in maniera casuale dai 508 casi iniziali, 93 sono stati ritenuti elegibili per la valutazione della concordanza intra ed inter-osservatore per la valutazione delle caratteristiche radiologiche (rN+, rENE, ed agglomerati linfonodali) dopo aver escluso 7 casi con intervallo di tempo tra imaging e chirurgia superiore ad 8 settimane. I coefficienti Kappa erano 0.78, 0.79, e 1.0 per la valutazione inter-osservatore e 0.89, 0.94, e 0.88 per la valutazione intra-osservatore di rN+, rENE, ed agglomerati linfonodali, rispettivamente (Tabella 1).

**Table 1** Diagnostic performance of radiologic features in 483 eligible cases; intrarater and interrater concordance in 93 randomly selected cases; radiologic features in relation to pathologic extranodal extension in 257 patients with radiologic overt lymph nodes

Diagnostic performance of radiologic features (n = 483)						
Covariate	All patients	Patients receiving CT	Patients receiving MRI	P value (CT vs MRI)		
<b>rN+ vs pN+</b>						
True Positive	195	122	73	-		
True Negative	159	107	52	-		
False Positive	62	44	18	-		
False Negative	67	34	33	-		
Sensitivity	74% (69-80)	78% (71-84)	69% (59-78)	.101		
Specificity	72% (66-78)	71% (63-78)	74% (62-84)	.660		
PPV	76% (70-81)	73% (66-80)	80% (71-88)	.211		
NPV	70% (64-76)	76% (68-83)	61% (50-72)	.017		
Accuracy	73%	75%	71%	.337		
<b>rENE+ vs pENE+</b>						
True Positive	51	34	17	-		
True Negative	92	63	29	-		
False Positive	4	3	1	-		
False Negative	48	22	26	-		
Sensitivity	52% (41-62)	61% (47-74)	40% (25-56)	.038		
Specificity	96% (90-99)	95% (87-99)	97% (83-100)	.660		
PPV	93% (82-98)	92% (78-98)	94% (73-100)	.787		
NPV	66% (57-74)	74% (63-83)	53% (39-66)	.010		
Accuracy	73%	80%	63%	<.001		
Intrarater and interrater reliability assessment of radiologic features						
Covariate	All patients		Patients receiving CT		Patients receiving MRI	
	n	$\kappa$ (95% CI)	n	$\kappa$ (95% CI)	n	$\kappa$ (95% CI)
<b>Intrarater reliability assessment</b>						
rN+	93	0.89 (0.80-0.98)	59	0.93 (0.84-1.00)	34	0.82 (0.63-1.00)
rENE	43	0.94 (0.83-1.00)	27	1.00	16	0.76 (0.32-1.00)
Matted LN	43	0.88 (0.64-1.00)	27	0.87 (0.62-1.00)	16	0.87 (0.61-1.00)
<b>Interrater reliability assessment</b>						
rN+	93	0.78 (0.66-0.91)	59	0.83 (0.69-0.97)	34	0.83 (0.69-0.97)
rENE	40	0.79 (0.57-1.00)	26	0.91 (0.73-1.00)	16	0.44 (0.28-1.00)
Matted LN	40	1.00	26	1.00	16	1.00
Various radiologic features in relation to pENE in 257 rN+ patients						
Covariate	rENE+ (ill-defined nodal border) (n = 55)		rENE- (n = 202)			
	pENE-Yes	pENE-No	pENE-Yes	pENE-No		
Total Case	51 (92%)	4 (8%)	48 (24%)	154 (76%)		
Invading adjacent structures	9 (100%)	0	0	0		
Matted nodes	11 (92%)	1 (8%)	0	0		
Surrounding edema	12 (100%)	0	0	0		
Nodal necrosis	35 (92%)	3 (8%)	4 (24%)	13 (76%)		
Logistic regression for pENE in pN+ cases						
		OR (95% CI)		P value		
rENE (yes vs no)		39.7 (9.8-152.0)		<.001		
Lymph node necrosis (yes vs no)		0.5 (0.2-1.6)		.240		

Abbreviations: CT = computed tomography; LN = lymph node; MRI = magnetic resonance imaging; NPV = negative predictive value; pENE = pathologic extranodal extension; PPV = positive predictive value; rENE = radiologic extranodal extension; rN+ = radiologic involved lymph node(s).

### Accuratezza diagnostica del rND versus il pND e del rENE versus pENE.

Le caratteristiche cliniche dei 483 casi finali sono descritte in Tabella 2. L'età media era 62 anni (range 21-92 anni). La più frequente sottosede di malattia era la lingua (227, 47%) seguita dal pavimento del cavo orale (95, 20%). La presenza di rN+ e rENE è stata valutata tramite TC (n= 307) o MRI (n=176). In 257 (53%) casi è stato evidenziato rN+, mentre 262 (54%) erano pN+. La sensibilità, specificità, PPV e NPV per rN+ predittivo di pN+ erano, rispettivamente, 74%, 72%, 76%, e 70% (Tabella 1).

**Table 2** Clinical characteristics of 483 oral cavity cancer cases

Covariate	Case Number (%)			P value
	All patients (n = 483)	Patients receiving CT (n = 307)	Patients receiving MRI (n = 176)	
Age median (range)	63 (21.8-92.6)	64 (24-93)	62 (22-87)	.013
Sex				.024
Female	174 (36)	99 (32)	75 (43)	
Male	309 (64)	208 (68)	101 (57)	
Subsite				<.001
Tongue	227 (47)	88 (29)	139 (79)	
Floor of mouth	65 (20)	77 (25)	18(10)	
Buccal mucosa	55 (11)	45 (15)	9(5)	
Lower alveolus and gingiva	54 (11)	51 (17)	3(2)	
Retromolar trigone	30 (6)	28 (9)	2(1)	
Upper alveolus and gingiva	11 (2)	9 (3)	2(1)	
Other (hard palate, mucosal lower lip)	12 (2)	9 (3)	3(2)	
Clinical T category				<.001
T1	109 (23)	69 (22)	40 (23)	
T2	187 (39)	109 (36)	78 (44)	
T3	62 (13)	30 (10)	32 (18)	
T4	125 (26)	99 (32)	26 (15)	
Clinical N category				.089
N0	221 (46)	151 (49)	70 (40)	
N1	86 (18)	47 (15)	39 (22)	
N2	175 (36)	108 (35)	67 (38)	
N3	1 (0)	1 (0)	0 (0)	
Clinical stage				.029
I	67 (14)	44 (14)	23 (13)	
II	89 (18)	52 (17)	37 (21)	
III	83 (17)	43 (14)	40 (23)	
IV	244 (51)	168 (55)	76 (43)	
pN+				.047
No	221 (46)	151 (49)	70 (40)	
Yes	262 (54)	156 (51)	106 (60)	
pENE+				.370
No	148 (56)	92 (59)	56 (53)	
Yes	114 (44)	64 (41)	50 (47)	
rN+				.640
No	226 (47)	141 (46)	85 (48)	
Yes	257 (53)	166 (54)	91 (52)	
rENE+ (among rN+)				.750
No rENE	202 (79)	129 (78)	73 (80)	
Yes rENE	55 (21)	37 (22)	18 (20)	
Radiologic matted node (among rN+)				.037
No	244 (95)	154 (93)	90 (99)	
Yes	13 (5)	12 (7)	1 (1)	
Radiologic invading adjacent structure (among rN+)				.120
No	245 (95)	161 (97)	84 (92)	
Yes	12 (5)	5 (3)	7 (8)	

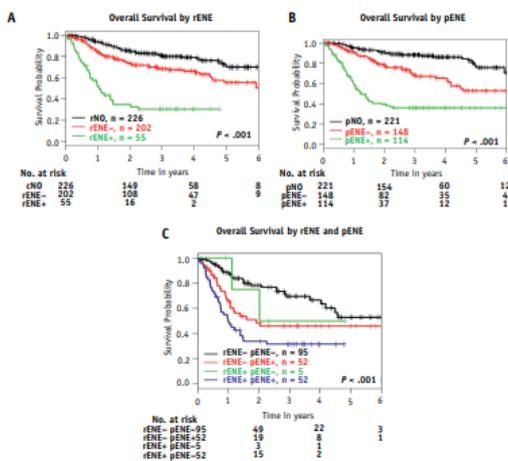
Abbreviations: CT = computed tomography; MRI = magnetic resonance imaging; pENE = pathologic extranodal extension; rENE = radiologic extranodal extension; rN+ = radiologic involved lymph node(s).

De i 257 pazienti con rN+, 55 erano rENE+. pENE è stata rilevata in 99 casi. La sensibilità, specificità, PPV e NPV del rENE nella rilevazione di pENE erano 52%, 96%, 93%, e 66%, rispettivamente. Dei 55 casi di rENE, 51 erano caratterizzati da pENE (OR 24.4, 95% CI: 8.3-71.2). La frequenza delle ulteriori caratterizzazioni radiologiche, tra cui gli agglomerati linfonodali (n =12), l'invasione delle strutture adiacenti (n=9), l'edema circostante (n =12), e la necrosi linfonodale (n = 55), in relazione alla presenza di rENE e pENE sono descritte in Tabella 1. La presenza di necrosi linfonodale è inoltre associata alla presenza di pENE (OR 3.5, 95% CI: 1.8-7.0). Tuttavia, essa diventa non predittiva (OR 0.50, 95% CI: 0.2-1.6, P = .24) dopo aver adeguato per rENE (OR 38.7, 95% CI: 9.8-152.0, P <.001) (Tabella 1). CT e MRI pre-operatorie erano disponibili in 307 e 176 casi, rispettivamente (nessuno dei paziente le aveva entrambe). La proporzione di pENE+ era simile tra CT e MRI (41% vs 47%, P = .370) benchè il tasso di pN+ era più alto con la MRI (60% vs 51%, P =

.047) (Tabella 2). Rispetto alla MRI, la TC ha dimostrato migliori sensitività (61% vs 40%,  $P = .038$ ), NPV (74% vs 53%,  $P = .010$ ), ed accuratezza diagnostica (80% vs 63%,  $P = .010$ ), ma simile specificità (95% vs 97%,  $P = .660$ ) e PPV (92% vs 94%,  $P = .787$ ) per rENE versus pENE dopo aver adattato per intervallo di tempo tra imaging e chirurgia (Tabella 1).

**Valore prognostico di rENE e pENE per l'OS.**

Il follow-up mediano per i pazienti sopravvissuti era di 3.0 anni (range: 0.6-6.6 years). L'OS a tre anni era significativamente ridotta per i casi rN+ con rENE ( $n = 55$ ) versus senza rENE ( $n = 202$ ), ed entrambi erano più bassi rispetto ai casi rN0 ( $n = 226$ ) (31% [20-47] versus 68% [62-76] versus 81% [75-87],  $P < .001$ ). L'OS a 3 anni per i casi pN+ con pENE versus senza pENE versus pN0 era 36% (28-47) versus 68% (60-78) versus 88% (84-93), rispettivamente ( $P < .001$ ). I casi caratterizzati sia da rENE+ che da pENE+ avevano la peggiore OS. L'OS a 3 anni per rENE+ e pENE+ (rENE+pENE+) ( $n = 52$ ) versus pENE+ ma rENE- (rENE-pENE+) ( $n = 52$ ) versus rENE- e pENE- (rENE-pENE-) ( $n = 95$ ) era 31% (21-48) versus 46% (33-63) versus 69% (59-81), rispettivamente ( $P < .001$ ). E' interessante notare che 5 casi erano caratterizzati da rENE senza pENE (rENE+pENE-) e la loro OS a 3 anni era 50% (19-100) (Fig. 3). La MVA ha confermato che sia rENE che pENE sono entrambi forti fattori prognostici per l'OS se adattati per età, T-, N-, e Zubrod PS. La hazard ratio (HR) per rENE e pENE era di 3.3 (2.0-5.3) e 2.6 (1.7-3.9), rispettivamente. rENE è rimasto ancora statisticamente significativo anco dopo l'aggiunta di pENE nel modello (HR 2.0,  $P = .015$ ) (Tabella 3)



Strata	Event/Total	3-Year OS	Strata	Event/Total	3-Year OS	Strata	Event/Total	3-Year OS
rN0	44 / 226	81% (75-87)	pN0	31 / 221	88% (84-93)	rENE- pENE-	28 / 95	69% (59-81)
rN+ rENE-	63 / 202	68% (62-76)	pN+ pENE-	45 / 148	68% (60-78)	rENE+ pENE-	25 / 52	46% (33-63)
rN+ rENE+	36 / 55	31% (20-47)	pN+ pENE+	67 / 114	36% (28-47)	rENE+ pENE+	2 / 5	50% (19-100)
<b>Total</b>	<b>143 / 483</b>	<b>70% (66-75)</b>	<b>Total</b>	<b>143 / 483</b>	<b>70% (66-75)</b>	<b>rENE+ pENE+</b>	<b>34 / 52</b>	<b>31% (21-48)</b>

**Fig. 3.** Overall survival in 483 oral cavity cancer cases stratified by radiologic or pathologic extranodal extension. (A) Overall survival stratified by radiologic extranodal extension. (B) Overall survival stratified by pathologic extranodal extension. (C) Overall survival by either radiologic or pathologic extranodal extension. Abbreviations: pENE+ = presence of pathologic extranodal extension; pENE- = absence of pathologic extranodal extension; pN+ = presence of pathologic involved lymph node; pN0 = absence of radiologic involved lymph node; rENE+ = presence of radiologic extranodal extension; rENE- = absence of radiologic extranodal extension; rN+ = presence of radiologic overt lymph node; rN0 = absence of radiologic overt lymph node.

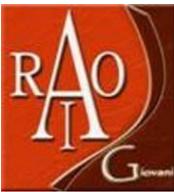
**Table 3** Multivariable analysis of OS by pENE and rENE

Model 1 — including rENE	HR (95% CI)	P value
rENE Yes vs no)	3.3 (2.4-5.3)	< .001
Age (10 years increment)	1.6 (1.3-1.9)	< .001
T category (7th) (T3-T4 vs T1-T2)	1.4 (0.9-2.2)	.150
N category (7th) (N2c-N3 vs N0-N2b)	2.2 (1.3-3.7)	.003
Zubrod PS ( $\geq 2$ vs 0-1)	1.8 (0.9-3.4)	.096
Model 2 — including pENE	HR (95% CI)	P value
pENE Yes vs no)	2.6 (1.7-3.9)	< .001
Age (10 years increment)	1.5 (1.3-1.8)	.120
T category (7th) (T3-T4 vs T1-T2)	1.4 (0.9-2.1)	.120
N category (7th) (N2c-N3 vs N0-N2b)	1.4 (0.9-2.3)	.140
Zubrod PS ( $\geq 2$ vs 0-1)	1.7 (0.9-3.0)	.086
Model 3 — including both rENE and pENE	HR (95% CI)	P value
pENE (Yes vs no)	2.0 (1.1-3.7)	.019
rENE (Yes vs no)	2.0 (1.1-3.4)	.015
Age (10 years increment)	1.6 (1.3-2.1)	< .001
T category (7th) (T3-T4 vs T1-T2)	1.5 (0.9-2.4)	.130
N category (7th) (N2c-N3 vs N0-N2b)	1.6 (1.0-2.8)	.076
Zubrod PS ( $\geq 2$ vs 0-1)	1.8 (0.9-3.6)	.090

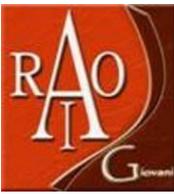
Abbreviations: CI = confidence interval; HR = hazard ratio; OS = overall survival; pENE = pathologic extranodal extension; PS = performance status; rENE = radiologic extranodal extension; rN+ = radiologic involved lymph node(s).  
C-index: Model 1: 0.71 (0.65-0.77); Model 2: 0.73 (0.67-0.79); Model 3: 0.74 (0.67-0.81).

**Discussione**

Questo ampio studio di coorte ha dimostrato che rENE è altamente specifico ma non sensibile per pENE, con una buona concordanza inter-osservatore ed un eccellente concordanza intra-osservatore. In maniera simile al pENE, la presenza di rENE è associata con una peggiore OS. Infine, La valutazione del rENE è più accurata se si utilizza la TC piuttosto che la MRI. Il valore prognostico di pENE per i tumori HNC non correlati a virus è stato dimostrato in diversi studi, ciò ha portato all'inclusione di questo parametro per la classificazione dell'N nell'ottava edizione del AJCC

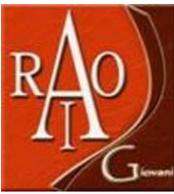


Cancer Staging Manual. In questa edizione la classificazione dell'N per i HNC non correlati a virus usa differenti definizioni per l'ENE clinico e patologico. Vista la mancanza di dati robusti circa l'accuratezza diagnostica nella determinazione dell'ENE con TC o MRI ed il suo incerto valore prognostico, la definizione del cENE nella classificazione dell'N dell'ottava edizione si riferisce a quei casi con evidenza certa di ENE clinica e radiologica. Secondo questa definizione, solo il 16% dei pazienti con rENE nello studio attuale (9 of 55) sarebbero stati classificati come pazienti con cENE. Mentre l'elevata specificità e valore prognostico del rENE (definito come "margini linfonodali mal-definiti") riportati in questo studio possono essere confermati in altri ampi datasets indipendenti e per altre sedi di malattia non-OSCC, criteri meno stringenti per cENE (es, usando margini linfonodali mal-definiti sulla TC o la MRI) potrebbero essere considerati in una revisione futura della classificazione di staging. Sebbene questo studio fosse indirizzato esclusivamente agli OSCC, è ipotizzabile che la performance diagnostica del rENE per il pENE potrebbe essere riproducibile per altri siti di HNC non correlato a virus. Tuttavia, dobbiamo essere cauti nell'estrapolare i risultati di questo studio ed applicarli negli HNC virus-relati poiché il microambiente tumorale potrebbe essere diverso in queste due tipologie di HNC. Ipoteticamente, questo tipo di caratteristica radiologica potrebbe difficilmente differenziabile da una estesa risposta infiammatoria circostante, che entrerebbe quindi in diagnosi differenziale con l'ENE. Uri et al, hanno valutato immagini TC di 49 casi di HNC, inclusi orofaringe (n = 16), laring (n = 13), cavo orale (n = 9), ipofaringe (n = 5), ed altre sedi di HNC (n = 6), riportando una sensibilità del 73% e una specificità del 91% per rENE versus pENE. Prabhu et al, hanno studiato l'accuratezza della TC in 432 casi di HNC, inclusi 325 OSCC e 107 tumori della laringe, riportando risultati simili (rENE vs pENE: sensibilità 43.7%; specificità 97.7%). Gli autori hanno inoltre dimostrato che la sensibilità del rENE migliora al crescere del grado di pENE (dal 19% al 72% per i gradi 1-2 vs il grado 3 di pENE). La difficoltà nell'identificare la pENE microscopica all'imaging non è una sorpresa, in quanto i cambiamenti radiologici potrebbero essere troppo impercettibili per essere rilevati ad occhio nudo, ciò potrebbe spiegare la bassa sensibilità del rENE versus pENE in questo studio. Nonostante in questo studio non vi fosse il grado di pENE a conferma, la più solida HR del rENE per l'OS comparata con pENE nella MVA supporta, in maniera indiretta, l'ipotesi secondo cui rENE possa identificare maggiori gradi di pENE. Nemmeno l'utilizzo della MRI sembra migliorare la sensibilità per quanto riguarda il rilievo di pENE microscopica (80% vs 20% tasso di falsi negativi per pENE microscopico vs macroscopico in uno studio). Che l'analisi radiomica possa essere di aiuto in questo ambito deve ancora essere valutato. Numerosi studi hanno esaminato la correlazione rENE-pENE; tuttavia, l'influenza dell'intervallo di tempo tra l'imaging preoperatorio e la chirurgia sull'accuratezza del rENE è sottovalutata. Sembra essere intuitivo che, data la natura aggressiva dell'OSCC, pENE potrebbe verificarsi nell'intervallo di tempo tra l'imaging preoperatorio e la chirurgia. Una progressiva riduzione dell'accuratezza nell'imaging preoperatorio dopo un intervallo >8 settimane suggerisce che potrebbe essere necessario effettuare un nuovo imaging nelle situazioni in cui l'intervallo di tempo con la chirurgia sia prolungato al fine di ottenere un staging adeguato. Nella pratica clinica, è importante comprendere l'anomala influenza di questo intervallo di tempo sull'efficienza della diagnostica per immagini dal momento che questo potrebbe influenzare lo standard di cura nella valutazione dei pazienti. Dai dati radiobiologici pubblicati, si può comprendere come il raddoppiamento del tumore possa aumentare durante i tempi di attesa. L'analisi iniziale presente nello studio circa l'influenza del tempo rispecchia il suddetto impatto. Il valore comparativo della MRI versus TC nella rilevazione della rENE nei pazienti affetti da HNC non è stato esplorato in maniera approfondita. Uno studio del 1992 effettuato da Yousem et al dimostrava una sensibilità massima dal 61% al 83% con immagini MRI T1- e T2-pesate con mezzo di contrasto. La limitazione principale di questo studio è stato l'affidamento degli autori sulla TC piuttosto che su l'esame istopatologico nella determinazione dell'ENE. King et al non hanno dimostrato alcuna differenza in sensibilità e specificità tra TC e MRI



in un piccolo studio composto da 17 pazienti. Nella analisi descritta nell'articolo, la TC ha mostrato di avere una migliore accuratezza per rENE se confrontata con la MRI(80% vs 63%,  $P= .011$ ). Mentre la MRI è solitamente migliore della TC nella delineazione dell'infiltrazione dei piani di tessuto grasso adiacenti e delle irregolarità dei contorni, questo vantaggio potrebbe essere oscurato dalla potenziale compromissione nella risoluzione correlata al movimento del paziente (es, respirazione, deglutizione) a causa del prolungato tempo di acquisizione di tale imaging rispetto a quello necessario per un esame TC. Inoltre, la MRI potrebbe essere soggetta ad una ricezione più debole del segnale al livello cervicale distale che potrebbe anch'essa ridurre la qualità dell'immagine. La valutazione del rENE sulla MRI nei casi dello studio è stata basata su immagini convenzionali T1 e T2-pesate, ma non è noto se altre sequenze MRI speciali (es, costante di assorbimento, immagini pesate in diffusione) possano aumentare l'accuratezza diagnostica. Ciononostante, la specificità del rENE sulla TC e sulla MRI appare simile (95% vs 97%), ciò supporta la combinazione della valutazione del rENE sia sulla TC che sulla MRI per la valutazione del valore prognostico del rENE. Alcuni dati emergenti hanno inoltre dimostrato rENE ha un valore prognostico per i HNC non virus-relati. Liu et al hanno analizzato 258 HNC localmente avanzati, inclusi orofaringe (n=83), unknown primary (n=13), cavo orale (n=11), e laringi (n=12) ed hanno mostrato una forte associazione tra un ridotto controllo di malattia ed un trend per ridotta OS nei HNC localmente avanzati con evidenza di rENE eccetto per i tumori dell'orofaringe HPV-correlati. Nello studio descritto in questo articolo, viene evidenziato come rENE sia un fattore prognostico indipendente per la OS per gli OSCC, aggiustando i valori per le altre covarianti. Questi risultati sottendono ad importanti implicazioni cliniche, specialmente quanto i dati patologici non sono disponibili (es., pazienti che non sono stati sottoposti a chirurgia). Se il valore prognostico del rENE potesse essere confermato nelle altre sottosedi di HNC non virus-relate (es. p16-negative and Epstein-Barr virus negative unknown primary, p16-negative orofaringe, ed ipofaringe p16-negative), ciò potrebbe comportare conseguenze dirette sulla stratificazione del rischio e sul piano di trattamento. Ad esempio, quando rENE viene evidenziato per le suddette sedi di malattia e gestito con approcci non chirurgici, potrebbe essere valutato un trattamento intensificato. Un'accurata identificazione preoperatoria dei pazienti con rENE sarebbe importante nei casi di tumore dell'orofaringe p16-negativi, per i quali una chirurgia robotica transorale potrebbe essere una potenziale alternativa alla radioterapia definitiva +/- chemioterapia. I pazienti con evidenza di rENE potrebbero essere trattati in maniera ottimale con CRT definitiva evitando l'ulteriore morbilità del trattamento trimodale.

L'alta specificità del rENE per il pENE influenzerebbe inoltre la delineazione del target nel piano di trattamento radioterapico, dove un più ampio margine del CTV è raccomandato quando un LN è associato ad ENE. Nell'articolo viene utilizzata la definizione "margini nodali mal-definiti" come unico criterio per definire rENE ed è stata evidenziata una buona concordanza inter-osservatore ed un'eccellente concordanza intra-operatore. Alcuni studi hanno, inoltre, incluso ulteriori caratteristiche radiologiche nella definizione del rENE (eg, invasione delle strutture circostanti, linfonodi confluenti, e/o disomogeneità nel grasso perinodale). Una meta-analisi condotta da Su et al, che valutava l'efficacia di differenti criteri diagnostici per rENE con la MRI, ha mostrato una specificità più alta usando sia "margini confusi" da sola (99%) o "contorno capsulare indistinto" da solo(93%) se confrontato con caratteristiche radiologiche multiple (margini linfonodali non distinti, indistinct nodal margins, enhancement irregolare della capsula linfonodale, o infiltrazione del grasso o dei muscoli adiacenti) (86%). Sulla base di questa esperienza, una definizione più generale sembra non necessaria dal momento che queste ulteriori caratteristiche radiologiche sono per lo più sempre accompagnate dalla presenza di margini linfonodali irregolari. E' interessante notare che, 1 caso che presentava linfonodi confluenti non aveva pENE. E' possibile che questo caso di "linfonodi agglomerati" sia stato diagnosticato erroneamente in quanto potrebbe essere più difficile definire i piani di clivaggio quando 3 linfonodi appaiono aggregati insieme piuttosto che nel singolo

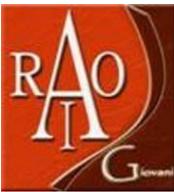


linfonodo. A causa della ridotta frequenza, non siamo in grado di identificare in maniera convincente altre caratteristiche radiologiche (eg, invasione delle strutture adiacenti e presenza di edema circostante), questo potrebbe essere potenzialmente utile nella ulteriore definizione del grado di rENE e nel migliorare l'accuratezza. E' plausibile che i progressi nelle tecniche di imaging, la radiomica o il machine learning possano ulteriormente migliorare la precisione del rilevamento del rENE e potenzialmente migliorare l'affidabilità inter-operatore. Le limitazioni di questo studio includono la natura retrospettiva e la scarsa distinzione tra pENE microscopica e macroscopica, poiché questa informazione non è compresa routinariamente in tutti i report patologici del loro istituto. Oltre a ciò, una piccola percentuale dei pazienti in esame non hanno effettuato la dissezione linfonodale cervicale o la dissezione di tutti i livelli. Tuttavia, si trattava di una percentuale ridotta (12%) del campione tale da influenzare difficilmente i risultati. Parallelamente, la differenza nella precisione diagnostica tra la TC e la MRI nel rilevamento del rENE è stata valutata in diversi sottogruppi dal momento che nessun paziente ha effettuato entrambi gli esami pre-operatori. Di conseguenza, non si può escludere un potenziale effetto confondente dovuto allo sbilanciamento di alcune caratteristiche. Ad esempio, l'estensione del pENE potrebbe essere differente anche se la proporzione di pENE+ era simile nello coorti della TC e della MRI. Infine, i risultati sono stati ottenuti da un registro monoistituzionale sull'OSCC. Che essi possano essere validati in altri datasets o generalizzati nelle altre sedi di HSC non virus-relate, deve essere ancora determinato. E' inoltre importante determinare se esista qualsiasi differenza apprezzabile nelle caratteristiche del rENE e nella concordanza rENE-pENE tra gli HNC virus-correlati versus non virus-correlati prima di generalizzare a tutti ai HNC virus-relati.

## **Conclusioni**

Questo ampio studio di coorte mono-istituzionale ha dimostrato che il moderno imaging (CT o MRI) è altamente specifico nella rilevazione del pENE nei pazienti affetti da OSCC, sebbene le relative sensibilità siano basse. La presenza di rENE è altamente indicativa di pENE ; tuttavia, l'assenza di rENE non è un buon indicatore per l'assenza di pENE. In aggiunta, la TC sembra essere la modalità di imaging da preferire rispetto alla MRI nella rilevazione del rENE. E' consigliabile eseguire gli esami diagnostici entro 8 settimane per definire con più accuratezza l'estensione di malattia, inclusa la presenza o l'assenza di rENE. Se si presentasse un ritardo nell'inizio del trattamento, sarebbe raccomandabile ripetere l'imaging. rENE è un fattore prognostico negativo indipendente per l'OS. Quando viene rilevato il rENE, si dovrebbe considerare un'intensificazione dell'approccio terapeutico. La bassa sensibilità del rENE non preclude il suo utilizzo come parametro per lo staging che fondamentalmente si basa su risultati positivi piuttosto che negativi per l'attribuzione (alta specificità con predittività forte); secondo le regole di stadiazione TNM, se c'è un dubbio (bassa sensibilità), è mandatario un approccio più conservativo, cioè, dovrebbe essere attribuito uno stadio meno negativo ad una caratteristica al fine di evitare il passaggio di stadio. Questo approccio è analogo nella valutazione clinica del collo durante la quale il clinico potrebbe rilevare evidente cENE secondo i criteri dell'ottava edizione del TNM come N3b con un'alta specificità, ma non può rilevare ENE microscopica che potrebbe, anch'essa, essere presente (bassa sensibilità).

Di conseguenza, nonostante la bassa sensibilità, l'alta specificità (forte certezza di ENE) e la sua importanza prognostica promuovono la considerazione di rENE nella stratificazione del rischio e nel piano di trattamento così come la sua potenziale inclusione in future revisione della classificazione clinica del parametro N.



1. Almulla A, Noel CW, Lu L, Xu W, O'Sullivan B, Goldstein DP, Hope A, Perez-Ordóñez B, Weinreb I, Irish J, Gullane P, Chepeha D, Tong L, Yu E, Huang SH. **Radiologic-Pathologic Correlation of Extranodal Extension in Patients With Squamous Cell Carcinoma of the Oral Cavity: Implications for Future Editions of the TNM Classification.** Int J Radiat Oncol Biol Phys. 2018 May 17. pii: S 0360-3016(18)30813-7. doi: 10.1016/j.ijrobp.2018.05.020. [Epub ahead of print] PubMed PMID: 29970315.