Supplementary material 7

Risk stratification models

Main study characteristics

Study (D or EV)	Model and cancer location	Source of data (n, and n included in analysis)	Age (mean y (SD)), sex	Cancer stage	TNM	HPV status and measurement method	Smoking, alcohol, co-morbidities	Treatment	Inclusion criteria/ comments on cohort representativeness.	Outcomes, length of follow-up
RTOG-0129 RPA risk stratification model										
Ang 2010, US *Development* *NB patient characteristics reported for n=323 with known HPV status*	RTOG-0129 RPA model Carcinoma of the oral cavity, oropharynx, hypopharynx, or larynx	Patients recruited July 2002 - May 2005 as part of randomised controlled trial; retrospective analysis; n=433; **n=266** (61%) in analysis	HPV+: median 53.5 (range 31-78); HPV-: 57 (387-82). 84% male.	TNM vr NR III: 14% IV: 86%	TNM vr NR T2: 31% T3: 39% T4: 30% N0: 7.4% N1: 14.2%, N2a:11.5% N2b: 33.1% N2c: 24.1% N3: 9.6%	HPV+: 64% HPV-: 36% Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV DNA detection (in situ hybridization)	HPV+: Median n of pack years 12.2 (range 0-152); HPV-: 36.5 (0-96). Never smoker (23%), former (51%), current (17%) and unknown (9%). No details on alcohol. Zubrod score 0-1: 71.9%, 2-3: 24.2%, unknown 3.8%.	Fractionation radiotherapy and cisplatin: 51% Accelerated-fractionation radiotherapy and cisplatin: 49%	US trial population (RTOG-129 study). Eligibility criteria: presence of untreated, pathologically confirmed, stage III or IV squamous-cell carcinoma of the oral cavity, oropharynx, hypopharynx, or larynx without distant	OS: defined as time from randomisation to death. PFS: defined as time from date of randomisation to death or first documented relapse categorised as local–regional recurrence or distant metastases. Median follow up 4.8 y (range, 0.3 to 6.5y)

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
									metastases (M0); Zubrod's performance status score of 0 or 1 (asymptomatic or symptomatic but ambulatory, respectively); age of 18 years or older; and adequate bone marrow, hepatic, and renal function.	
Granata 2012, Italy *External validation*	RTOG-0129 RPA model OPSCC (no further details)	Consecutive patients treated at one Italian institution 2003 - 2009; retrospective analysis; n=140; n= **120** (86%) in analysis	Median (range) 61.0 (37-88) 80% male	TNM vr NR I: 0.8% II: 5.8% III: 8.3% IVa:75.8 % IVb: 9.2%	TNM vr NR T1: 14.2% T2: 30.8% T3: 12.5% T4: 41.7%	p16+: 54.6% p16-:45.4% Tumour p16 protein expression (immunohistochemical analysis)	Median (range) packs/year: 30 (0-120). No smoking: 20.8%, smoking: 79.2% No details on alcohol.	Induction CT +CRT: 32.5% CRT: 53.3% RT only: 14.2%	Single Italian institution (consecutive patients). Not explicitly stated whether patients treated with	OS: defined as time from the date of starting therapy to the date of death due to any cause, censoring at the date of last available follow-up assessment for living patients. Median follow-up of 23 months (interquartile range: 13–39 months)

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
					Tis: 0.8% N0: 12.5% N1: 10% N2a: 2.5% N2b: 40.8% N2c: 25% N3: 9.2%		ECOG status-0: 73.3%, 1: 23.3%, 2: 3.4%		curative intent.	
Rietbergen 2013, The Netherlands *External validation*	RTOG-0129 RPA model OPSCC HPV+: tonsil 58.9%, base of tongue 31.3%, soft palate 5.5%, oropharynx nos 4.3%; HPV-: tonsil 38.4%, base of	All patients with an OPSCC diagnosed at two Dutch University hospitals Jan 2000 - Dec 2006; retrospective analysis; n=906; **n=721** in analysis (those treated with curative intent and with model	HPV+: mean age at diagnosis 60.53 (median 58.14), HPV-: mean 60.94 (median 59.67). 67% male *NB based on n=809*	TNM vr NR I-VI	TNM vr NR HPV+: T1-2: 54% T3-4: 46% Tx: 0 N0: 14.7% N1-3: 85.3% Nx: 0 HPV-: T1-2: 39.1% T3-4:	HPV+: 19.4% HPV-: 76.8% Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV DNA detection (PCR) *NB based on n=809*	HPV+: 0-10 pack years 40.5%, 11-24 PY 16.0%, >24 PY 42.9%, unknown 0.6%. HPV-: 0-10 PY 6.3%, 11-24 PY 9.3%, >24 PY 83.1%, unknown 1.2%. HPV+: 0-100 unit yrs 68.7%, 111-149 UY 9.2%, >149 UY 21.5%, unknown 0.6%. HPV-: 0-100	HPV+: Curative 93.3%, palliative 6.7%, Surg +/- RT 27.7% RT 18.4%, CRT 27.6%, RT+LND+ RT (brachytherapy) 26.3%; HPV- curative 84.4%,	Unselected Dutch cohort of patients with OPSCC treated with curative intent.	OS: defined as the time from the date of incidence (defined as the date on which the squamous cell carcinoma was histologically confirmed) to death (any cause). PFS: defined as the time period from date of incidence to death or the first documented relapse, which was categorized as local-regional recurrence or distant metastases. Median follow-up of patients who received treatment and remained alive was 4.33 years (range 0.1–12.1)

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
	tongue 24.9%, soft palate 17.8%, oropharynx nos 18.9%.	parameter data).			60.9% Tx: 4.0% N0: 39.3% N1-3: 60.4% Nx: 0.3%		UY 25.9%, 111-149 UY 7.7%, >149 UY 64.2%, unknown 2.2% HPV+: ACE-27 score 0: 56.4%, 1:24.5%, 3:17.2%, unknown 0.6%. HPV-: 0:32.4%, 1:30.3%, 2:28.8%, 3:8.4%, unknown: 0.2%	palliative 15.5%, surg +/-RT 30.7%, RT 31%, CRT 28.8%, RT +LND+RT (brachytherapy): 9.5%		
Rietbergen 2015, The Netherlands *External validation*	RTOG-0129 RPA model OPSCC (no further details)	Patients curatively treated 2000–2011 at Maastricht University Medical Center, The Netherlands; retrospective analysis; **n=235**	Mean 60.24, median 58.96 73.6% male	TNM vr NR I-II: 20.4% III-IV: 79.6%	TNM vr NR T1-2: 51.5% T3-4: 48.5% Tx=0 N0-N2a: 48.1% N2b-N3: 51.9% Nx:0	HPV+: 30.2% HPV-: 69.8% Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV DNA detection (PCR)	0-10 PY: 18.7%, >10 PY: 81.3%) unknown: 0 Details collected but not reported. ACE score 0-1: 71.9%, 2-3: 24.2%, unknown 3.8%.	Surgery +/- RT: 30.2% RT: 46% CRT: 19.1% Other: 4.7%	Single Dutch institution. Patients treated with curative intent.	OS: defined as time from date of incidence (defined as the date on which the squamous cell carcinoma was histologically confirmed) to death (any cause). PFS: defined as time period from date of incidence to death or the first documented relapse that was categorised as local–regional recurrence or distant metastases. Mean/median follow-up time not stated. Up to 5 years.

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
Wang 2016, Taiwan *External validation*	RTOG-0129 RPA model OPSCC stage III-IV (57.5% tonsil, 23.9% tongue base, 10.6% pharyngeal wall, 8.0% soft palate)	Patients with non-metastatic stage III–IV OPSCC who had completed definite therapy June 2006 -Jan 2013 at one hospital; retrospective analysis; n=130; **n=113** (87%) in analysis.	Range 28-86, 52.2% >50 years 90.3% male	TNM 7th ed III: 10.6% IVa: 64.6% IVb: 24.8%	TNM 7th ed T1-T3: 44.2% T4:55.8 % N0-N2a: 23.0%, N2b-N3: 77.0%	P16+: 24.8% P16-: 75.2% Tumour p16 protein expression (immunohistochemical analysis)	72.6% smokers No details on alcohol. No details on co-morbidity.	Cisplatin-based CCRT: 75.2%, induction chemotherapy and CCRT: 18.6%, cetuximab bio-radiotherapy: 4.4%, radiotherapy alone: 1.8%.	Single Taiwanese hospital. Patients who had cancer and achieved complete remission for more than three years without CT or RT were eligible for the current study. Patients were excluded if they had been previously treated for OPSCC at other institutions or had a positive history of malignancies for which they had	OS: not defined. DSS: not defined. At least 24 months or until death. Median follow-up time 27.6 months (range: 5.2–98.0 months).

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
									been treated with CT or RT.	
Fakhry 2017, US *Development cohort (for Fakhry model), used for external validation of RTOG-0129 model*	RTOG-0129 RPA model OPSCC (no further details)	Patients from RTOG 0129 and 0522 clinical trials (2002-2005 and 2005-2009), not consecutive, **n=493**	Mean age NR; 27.6% ≤50 yrs, 72.4% >50 yrs 87.4% male	All Stage III-IV as per the original RTOG studies.	TNM 7th and 8th eds T2-3: 72.0% T4: 28.0% N0-2b (TNM7) or N0-N1 (TNM8): 66.5% N2c-3 (TNM7) or N2-N3 (TNM8): 33.5%	P16 +: 73.6% P16-: 26.4% Tumour p16 protein expression (immunohistochemical analysis)	≤10 PYs: 47.3%, >10 PYs: 52.7% Not reported Zubrod/ECOG/ WHO PS: PS0: 69.0%; PS1: 31.0%	All patients curative RT+CT.	Patients from two US trials (NRG Oncology RTOG 0129 and 0522). Eligible patients had untreated, pathologically confirmed, AJCC 5th edition (RTOG 0129) or 6th edition (RTOG 0522) stage III to IV7 head and neck squamous cell carcinoma, Zubrod performanc	OS: defined as time from date of randomisation to death from any cause. PFS: defined as from date of randomization to local, regional, or distant progression or death from any cause. Median: 5.7 years (95% CI, 5.2 to 6.0 years)

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
									e status of 0 to 1, age ≥ 18 years, and adequate bone marrow, hepatic, and renal function.	
Deschuymer 2018, Belgium *External validation*	RTOG-0129 RPA model OPSCC (no further details)	All non-metastatic OPSCC patients treated with primary RT in a single institution Jan 2004 -July 2017; retrospective analysis; n=333; **n=258** (77%) in analysis.	HPV+ mean (IQR) 63.6 (54.9-70.1); HPV- 59.7 (54.0-65.8). 79.2% male.	TNM 7th ed HPV+: I: 0.% II:10.0% III: 11.0% IVa: 68.0% IVb: 11.0% HPV-: I: 0.63% II:11.88% III:16.25 % IVa: 57.5% IVb: 13.75%	TNM 7th ed HPV+: T1: 12.0% T2: 41.0% T3: 19.0% T4a: 23.0% T4b: 5.0% N0: 17.0% N1: 10.0% N2a: 5.0% N2b: 42.0%,	HPV+: 38.5% HPV-: 61.5% Tumour p16 protein expression (immunohistochemical analysis)	HPV+: mean PY 22.6 (21.65), never smoker 27.0%, ≤10 PY 14.0%, >10 PY 57.0%, unknown 3.0%. HPV-: mean PY 38.8 (22.12), never smoker 1.88%, ≤10 PY 5.0%, >10 PY 91.87%, unknown 1.25%. No details on alcohol. HPV+: ACE27 0 : 31.0%, 1: 37.0%, 2: 24.0%, 3:8.0%.	All primary radiotherapy. HPV+ no systematic treatment 31.0%, cisplatin 66.0%, EGFR inhibitor 3.0%. HPV- no systematic treatment 28.75%, cisplatin 60.63%, EGFR inhibitor 17.0%.	Single Belgian institution. Included only patients treated with curative primary RT, excluding stage IV (8th Ed) and excluding patients treated with primary surgery.	OS: calculated from the date of histological diagnosis to the date of death from any cause. LRC: calculated from the date of histological diagnosis to the date of locoregional relapse (tumour at the primary site or regional nodes). DMC: calculated from the date of histological diagnosis to the date of distant metastases. Median follow-up 63.7 months (IQR 30.0; 99.9).

Study (D or EV)	Model and cancer location	Source of data (n, and n included in analysis)	Age (mean y (SD)), sex	Cancer stage	TNM	HPV status and measurement method	Smoking, alcohol, co-morbidities	Treatment	Inclusion criteria/ comments on cohort representativeness.	Outcomes, length of follow-up
					N2c:20.0% N3: 6.0%. HPV-: T1: 0% T2: 33.13% T3: 22.5% T4a: 26.25% T4b:11.88% N0: 23.13% N1: 17.5% N2a: 2.5% N2b: 26.88% N2c: 28.13% N3: 1.88%		HPV-: ACE270: 16.25%, 1:40.0%, 2:23.75%, 3:20.0%.			

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
Rios-Velaquez 2014, The Netherlands *External validation Cohort 1*	RTOG-0129 RPA model OPSCC (Tonsillar fossa (36.9%), base of tongue (29.8%), oropharynx overlap (25.6%), soft palate 7.7%))	Consecutively treated patients (Maastro Clinic) Jan 2000-Oct 2011; retrospective analysis; **n=168**	Median 59 (range 43-83) 74.4% male	TNM vr NR I–IVb (no further details)	TNM vr NR T1: 14.9% T2: 27.4% T3: 22.6% T4: 35.1% N0: 34.5% N1: 17.3% N2: 44.1% N3: 3.6% Nx: 0.6%	P16+: 34.5% P16- : 64.3% P16 unknown: 1.25% HPV_DNA+:30.4% HPV_DNA-: 69.6% Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV-16 DNA detection (PCR)	Median 30PY (range 0-100) Median 134 units/years (range 0-660) ACE 27 None: 33.3% Mild: 41.1% Moderate: 19% Severe: 6.5%	RT only: 67.9% CRT: 32.1%	Single Dutch institution. Patients treated with curative intent.	OS: defined as the time from starting radiotherapy to death from any cause. PFS: defined as the time from starting radiotherapy to time of first documented recurrence at any site (locoregional or metastasis) or death from any cause. Median 26 months (range 2.5–127.2) overall. Median 37.5 months (range 6.4–127.2) for patients alive at last follow-up.
Rios-Velaquez 2014, The Netherlands *External validation Cohort 2*	RTOG-0129 RPA model OPSCC (Tonsillar fossa (38.1%), base of tongue (29.1%), oropharynx	Consecutively treated patients (VU University Medical Center) Jan 2000-Dec 2006; retrospective analysis; **n=189**	Median 60 (range 43– 93) 64.6% male	TNM vr NR I–IVb (no further details)	TNM vr NR T1: 13.8% T2: 28% T3: 33.3% T4: 24.9% N0:	P16+: 16.9% P16- : 82% P16 unknown: 1.1% HPV DNA+: 18% HPV DNA-: 82% Tumour p16 protein	Median 32 PY (range 0-100) Median 170 units/years (range 0-350) ACE 27 None: 35.4% Mild: 29.1%	RT only: 60.8% CRT: 39.2%	Single Dutch institution. Patients treated with curative intent.	OS: defined as the time from starting radiotherapy to death from any cause. PFS: defined as the time from starting radiotherapy to time of first documented recurrence at any site (locoregional or metastasis) or death from any cause.

Study (D or EV)	Model and cancer location	Source of data (n, and n included in analysis)	Age (mean y (SD)), sex	Cancer stage	TNM	HPV status and measurement method	Smoking, alcohol, co-morbidities	Treatment	Inclusion criteria/ comments on cohort representativeness.	Outcomes, length of follow-up
	overlap (20.1%), soft palate 12.7%))				43.9% N1: 12.2% N2: 41.8% N3: 2.1% Nx: 0%	expression (immunohistochemical analysis); Tumour HPV-16 DNA detection (PCR)	Moderate: 29.6% Severe: 5.8%			Not reported.
Rietbergen 2013 risk stratification model										
Rietbergen 2013, The Netherlands *Development*	Rietbergen 2013 model OPSCC HPV+: tonsil 58.9%, base of tongue 31.3%, soft palate 5.5%, oropharynx nos 4.3%; HPV-: tonsil	All patients with an OPSCC diagnosed at two Dutch University hospitals Jan 2000 - Dec 2006; retrospective analysis; n=906; **n=721** in analysis (those treated with curative	HPV+: mean age at diagnosis 60.53 (median 58.14), HPV-: mean 60.94 (median 59.67). 67% male	TNM vr NR I-IV	TNM vr NR HPV+: T1-2: 54% T3-4: 46% Tx: 0 N0: 14.7% N1-3: 85.3% Nx: 0 HPV-:	HPV+: 19.4% HPV-: 76.8% *NB based on n=809* Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV DNA detection (PCR)	HPV+: 0-10 pack years 40.5%, 11-24 PY 16.0%, >24 PY 42.9%, unknown 0.6%. HPV-: 0-10 PY 6.3%, 11-24 PY 9.3%, >24 PY 83.1%, unknown 1.2%. HPV+: 0-100 unit yrs 68.7%, 111-149 UY 9.2%, >149 UY 21.5%,	HPV+: Curative 93.3%, palliative 6.7%, Surg +/- RT 27.7%, RT 18.4%, CRT 27.6%, RT+LND+ RT (brachytherapy) 26.3%;	Unselected Dutch cohort of patients with OPSCC treated with curative intent.	OS: defined as the time from the date of incidence (defined as the date on which the squamous cell carcinoma was histologically confirmed) to death (any cause). PFS: defined as the time period from date of incidence to death or the first documented relapse, which was categorized as local-regional recurrence or distant metastases. Median follow-up of patients who received treatment and remained alive was 4.33 years (range 0.1–12.1)

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
	38.4%, base of tongue 24.9%, soft palate 17.8%, oropharynx nos 18.9%.	intent and with model parameter data).	*NB based on n=809*		T1-2: 39.1% T3-4: 60.9% Tx: 4.0% N0: 39.3% N1-3: 60.4% Nx: 0.3%		unknown 0.6%. HPV-: 0-100 UY 25.9%, 111-149 UY 7.7%, >149 UY 64.2%, unknown 2.2% HPV+: ACE-27 score 0: 56.4%, 1:24.5%, 3:17.2%, unknown 0.6%. HPV-: 0:32.4%, 1:30.3%, 2:28.8%, 3:8.4%, unknown: 0.2%	HPV-curative 84.4%, palliative 15.5%, surg +/-RT 30.7%, RT 31%, CRT 28.8%, RT +LND+RT (brachytherapy): 9.5%		
Rietbergen 2015, The Netherlands *External validation*	Rietbergen 2013 model OPSCC (no further details)	Patients curatively treated 2000–2011 at Maastricht University Medical Center; retrospective analysis; **n=235**	Mean 60.24, median 58.96 73.6% male	TNM vr NR I-II: 20.4% III-IV: 79.6%	TNM vr NR T1-2: 51.5% T3-4: 48.5% Tx=0 N0-N2a: 48.1%	HPV+: 30.2% HPV-: 69.8% Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV DNA detection (PCR)	0-10 PY: 18.7%, >10 PY: 81.3%) unknown: 0 Details collected but not reported. ACE score 0-1: 71.9%, 2-3: 24.2%, unknown 3.8%.	Surgery + RT: 30.2%; RT: 46%; CRT: 19.1%; other: 4.7%	Single Dutch institution. Patients treated with curative intent.	OS: defined as time from date of incidence (defined as the date on which the squamous cell carcinoma was histologically confirmed) to death (any cause). PFS: defined as time period from date of incidence to death or the first documented relapse that was categorised as local–regional recurrence or distant metastases.

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
					N2b-N3: 51.9% Nx:0					Mean/median follow-up time not stated. Up to 5 years.
Deschuymer 2018, Belgium *External validation*	RTOG-0129 RPA model OPSCC (no further details)	All non-metastatic OPSCC patients treated with primary RT in a single institution Jan 2004 -July 2017; retrospective analysis; n=333; **n=258** (77%) in analysis.	HPV+ mean (IQR) 63.6 (54.9-70.1); HPV- 59.7 (54.0-65.8). 79.2% male.	TNM 7th ed HPV+: I: 0.% II:10.0% III: 11.0% IVa: 68.0% IVb: 11.0% HPV-: I: 0.63% II:11.88% III:16.25 % IVa: 57.5% IVb: 13.75%	TNM 7th ed HPV+: T1: 12.0% T2: 41.0% T3: 19.0% T4a: 23.0% T4b: 5.0% N0: 17.0% N1: 10.0% N2a: 5.0% N2b: 42.0%, N2c:20.0% N3: 6.0%.	HPV+: 38.5% HPV-: 61.5% Tumour p16 protein expression (immunohistochemical analysis)	HPV+: mean PY 22.6 (21.65), never smoker 27.0%, ≤10 PY 14.0%, >10 PY 57.0%, unknown 3.0%. HPV-: mean PY 38.8 (22.12), never smoker 1.88%, ≤10 PY 5.0%, >10 PY 91.87%, unknown 1.25%. No details on alcohol. HPV+: ACE27 0 : 31.0%, 1: 37.0%, 2: 24.0%, 3:8.0%. HPV-: ACE27 0: 16.25%, 1:40.0%, 2:23.75%, 3:20.0%.	All primary RT. HPV+ no systematic treatment 31.0%, cisplatin 66.0%, EGFR inhibitor 3.0%. HPV- no systematic treatment 28.75%, cisplatin 60.63%, EGFR inhibitor 17.0%.	Single Belgian institution. Included only patients treated with curative primary RT, excluding stage IV (8th Ed) and excluding patients treated with primary surgery.	OS: calculated from the date of histological diagnosis to the date of death from any cause. LRC: calculated from the date of histological diagnosis to the date of locoregional relapse (tumour at the primary site or regional nodes). DMC: calculated from the date of histological diagnosis to the date of distant metastases. Median follow-up 63.7 months (IQR 30.0; 99.9).

Study (D or EV)	Model and cancer location	Source of data (n, and n included in analysis)	Age (mean y (SD)), sex	Cancer stage	TNM	HPV status and measurement method	Smoking, alcohol, co-morbidities	Treatment	Inclusion criteria/ comments on cohort representativeness.	Outcomes, length of follow-up
					HPV-: T1: 0% T2: 33.13% T3: 22.5% T4a: 26.25% T4b:11.88% N0: 23.13% N1: 17.5% N2a: 2.5% N2b: 26.88% N2c: 28.13% N3: 1.88%					
Huang 2015 RPA and AHR risk stratification models										
Huang 2015, Canada *Development*	Huang 2015 RPA model and AHR model	Consecutive patients treated 2000-2010 at one institution; retrospective analysis; **n=573**	Median 57.8 79% male	TNM 7th ed I: 1.4% II: 4.4% III: 13.8% IV: 80.5%	TNM 7th ed T1: 20.4% T2: 34.6%	100% HPV + (p16 +/- staining) Tumour p16 protein expression (immunohistoc	Median 40 PYs Not reported Not reported	All radiotherapy CRT: 50%	Unselected patient population with HPV-related OPC from a single Canadian	OS: defined as the duration between “random assignment” and death. Median 5·1 years

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
	OPSSC (HPV+ only)	NB based on 1,108 patients, 573HPV+, 237 HPV- and 298 (27%) with unknown status			T3: 26.9 % T4: 18.2% N0: 12% N1: 10.5% N2a: 7.9% N2b: 36.3% N2c: 24.1% N3: 9.3%	hemical analysis)			institution treating almost all patients in its region. Patients with nonmetastatic (M0) OPC treated with definitive RT or CRT.	
Keane 2016, US *External validation*		Patients from SEER database, diagnosed 2004-2008; retrospective analysis; **n=8427**	Median 58 (21–96) 82.8% male.	TNM 6th ed I: 5.9% II: 8.9% III: 22.6% IVA: 57.5% IVB: 5.2%	TNM 6th ed T1: 25.3% T2: 39.2% T3: 12.5% T4: 23.0% N0: 21.9% N1: 24.8% N2a: 10.6%	HPV+ (no further information on method of HPV status determination). Method not reported.	Not reported Not reported Not reported	Radiotherapy No: 14.4% Yes: 85.6% Radical surgery No: 72.1 Yes: 27.7% Unknown: 0.02%. No information on CT.	Patients with non-metastatic cancer from US SEER database. No information on CT, smoking, alcohol or co-morbidities. Unclear if cohort different to DEV one.	OS: defined as time from date of diagnosis to date of death. HNC-specific mortality: defined as time from date of diagnosis to date of head and neck cancer related death. Median follow-up 50 months

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
					N2b: 24.7% N2c: 12.9% N3: 5.2% All M0				Not possible to determine whether patients received definitive-intent surgery.	
O'Sullivan 2016, Canada *External validation*		Consecutive patients treated 2000-2011 at six institutions; retrospective analysis; **n=1246** NB based on patients with known HPV+ status	Median 56 (IQR 51–62) 86% male	TNM 7th ed I: 1% I: 3% III: 13% IVA: 76% IVB: 7%	TNM 7th ed T1: 30% T2: 39% T3: 19% T4a: 11% T4b: 1% N0: 7% N1: 12% N2a: 12% N2b: 41% N2c: 22%	100% HPV + (p16 +/- staining) Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV DNA detection (in situ hybridization)	Median PYs 6 (IQR 0-25) Not reported Not reported	Surgery: 2% RT: 98% CT yes: 75% CT no: 25%	Patients with non-metastatic oropharyngeal cancer from six institutions, one from Denmark, one from the Netherlands and four from the US. Patients with newly diagnosed non-metastatic (M0)	Risk of death: defined as risk of death from any cause from the date of diagnosis. Median 4·6 years (IQR 3·1–5·5)

Study (D or EV)	Model and cancer location	Source of data (n, and n included in analysis)	Age (mean y (SD)), sex	Cancer stage	TNM	HPV status and measurement method	Smoking, alcohol, co-morbidities	Treatment	Inclusion criteria/ comments on cohort representativeness.	Outcomes, length of follow-up
					N3: 6% All M0				oropharyngeal cancer undergoing either primary surgery or primary radiotherapy with or without chemotherapy. No explicit statement on treatment with curative intent.	
O'Sullivan 2016 AHR and RPA risk stratification model (updated from Huang 2015)										
O'Sullivan 2016, Canada *Development*	O'Sullivan 2016 AHR model OPSSC (HPV+ only)	Consecutive patients treated 2000-2011 at one institution; retrospective analysis, **n=661** NB based on patients with known HPV+ status	Median 57 (IQR 51–65) 80% male	TNM 7th ed I: 1% II: 4% III: 13% IVA: 68% IVB: 13%	TNM 7th ed T1: 20% T2: 35% T3: 26% T4a: 13%	100% HPV + (p16 +/- staining) Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV DNA detection	Median PYs 15 (IQR 0-30) Not reported Not reported	Surgery: 1% RT: 99% CT yes: 49% CT no: 51%	Patients with non-metastatic oropharyngeal cancer from one Canadian institution. Patients with newly diagnosed non-	Risk of death: defined as risk of death from any cause from the date of diagnosis. Median 5·5 years (IQR 3·2–6·6)

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
					T4b: 5% N0: 12% N1: 10% N2a: 8% N2b: 36% N2c: 25% N3: 8% All M0	(in situ hybridisation).			metastatic (M0) oropharyngeal cancer undergoing either primary surgery or primary RT with or without CT.	
O'Sullivan 2016, Canada *External validation*		Consecutive patients treated 2000-2011 at six institutions; retrospective analysis; **n=1246** NB based on patients with known HPV+ status	Median 56 (IQR 51–62) 86% male	TNM 7th ed I: 1% I: 3% III: 13% IVA: 76% IVB: 7%	TNM 7th ed T1: 30% T2: 39% T3: 19% T4a: 11% T4b: 1% N0: 7% N1: 12% N2a: 12%	100% HPV + (p16 +/- staining) Tumour p16 protein expression (immunohistochemical analysis); Tumour HPV DNA detection (in situ hybridisation).	Median PYs 6 (IQR 0-25) Not reported Not reported	Surgery: 2% RT: 98% CT yes: 75% CT no: 25%	Patients with non-metastatic oropharyngeal cancer from six institutions, one from Denmark, one from the Netherlands and four from the US. Patients with newly diagnosed non-	Risk of death: defined as risk of death from any cause from the date of diagnosis. Median 4·6 years (IQR 3·1–5·5)

Study (D or EV)	Model and cancer location	Source of data (n, and n included in analysis)	Age (mean y (SD)), sex	Cancer stage	TNM	HPV status and measurement method	Smoking, alcohol, co-morbidities	Treatment	Inclusion criteria/ comments on cohort representativeness.	Outcomes, length of follow-up
					N2b: 41% N2c: 22% N3: 6% All M0				metastatic (M0) oropharyngeal cancer undergoing either primary surgery or primary RT with or without CT.	
Alabi 2022 Risk stratification model										
Alabi 2022 *Development*	ProgTOOL OPSCC (Tonsil 55.1%), base of tongue (37.2%), oropharynx (6.9%), vallecular 0.8%))	Patients from SEER database 2010-2015; retrospective analysis; **n=3164** Patients with known clinical and pathological characteristics.	Median 61 (SD 10.4, range 20-85; mean 61.4) 79.8% male	Not reported	TNM 7th ed T1: 27.6% T2: 40.4% T3: 18.3% T4: 13.6% N0: 44.8% N1: 45% N2: 0%	HPV+: 63.9% HPV-: 36.1% No further information on method of HPV status determination.	Not reported Not reported Not reported	Surgery: 18.3% Surgery + RT: 40.1% Surgery + CRT: 18.3% CRT: 13.1% No treatment: 12.1%	Included all cases with OPSCC with known clinical and pathologic characteristics.	OS: defined as the time period from the beginning (or end) of treatment until the patients die of any cause. Median 49, mean 49.4, SD 27.2, range 0 to 107 months.

Study (D or EV)	**Model and cancer location**	**Source of data (n, and n included in analysis)**	**Age (mean y (SD)), sex**	**Cancer stage**	**TNM**	**HPV status and measurement method**	**Smoking, alcohol, co-morbidities**	**Treatment**	**Inclusion criteria/ comments on cohort representativeness.**	**Outcomes, length of follow-up**
					N3: 10.2% M0: 97.5% M1:2.5 %					
Alabi 2023 *External validation*	ProgTOOL OPSCC (Tonsil 58.9%), base of tongue (28.2%), oropharynx (12.9%).	Case of OPSCC from electronic patient records at Helsinki University Hospital; retrospective analysis; **n=163** Patients with known clinical and pathological characteristics.	Median 62 (SD 9.1, range 37-85, mean 61.4) 75.5% male	Not reported	TNM 7th ed T1: 32.5% T2: 32.5% T3: 10.4% T4: 24.5% N0: 15.3% N1: 9.8% N2: 73.6% N3: 1.2% M0: 98.8% M1:1.2 %	HPV+: 74.2% HPV-: 25.8% No further information on method of HPV status determination.	Not reported Not reported Not reported	Surgery: 4.9% Surgery + RT: 14.1% Surgery + CRT: 21.5% CRT: 51.5% No treatment: 8.0%	Included cases with OPSCC with known clinical and pathologic characteristics.	OS: not defined Median 50, mean 45.8, SD 19.9, range 0 to 88 months.

Model characteristics and performance

Ang 2010, US *Development*	**RTOG-0129 RPA model** HPV status: positive vs negative Number of pack-years of tobacco smoking: ≤10 vs >10 Nodal/tumour stage: N0 to N2a vs N2b to N3 for HPV-positive tumours OR: tumour stage (T2 or T3 vs T4), for HPV-negative tumours *Flowchart presented on how to assign low, intermediate and high risk.*	*Model discrimination (c-index)*: NR *Model calibration*: NR *Other model performance measures*: patients classified as having low, intermediate or high risk of 3-yr OS using RPA.
Granata 2012, Italy *External validation*		*Model discrimination (c-index)*: 2yr OS: c-index (adjusted for optimism) 0.70 (no CI stated) *Model calibration*: Observed event rate for OS in external validation set was slightly higher than the observed event rate for OS in the DEV cohort for all three predicted risk groups. *Other model performance measures*: NR
Rietbergen 2013, The Netherlands *External validation*		*Model discrimination (c-index)*: OS: c-index 0.58 (95% CI 0.56-0.61) *Model calibration*: NR *Other model performance measures*: NR
Rietbergen 2015, The Netherlands *External validation*		*Model discrimination (c-index)*: 5 yr OS: c-index 0.65 (95% CI 0.59, 0.70); 5 yr PFS: c-index 0.61 (95% CI 0.54, 0.68) *Model calibration*: NR *Other model performance measures*: NR
Wang 2016, Taiwan *External validation*		*Model discrimination (c-index)*: OS: c-index 0.67 (CI NR); DSS: c-index 0.68 (CI NR) *Model calibration*: NR *Other model performance measures*: NR
Fakhry 2017, US *External validation*		*Model discrimination (c-index)*: OS: c-index 0.71 (95% CI 0.66, 0.76) *Model calibration*: NR *Other model performance measures*: NR
Deschuymer 2018, Belgium *External validation*		*Model discrimination (c-index)*: OS: c-index 0.57 (95% CI 0.52, 0.62); OS HPV+ group only: c-index 0.62 (95% CI 0.51, 0.73); c-index not reported for LRC or DMD *Model calibration*: NR *Other model performance measures*: NR
Rios-Velazquez 2014 *External validation (Maastro cohort)*		*Model discrimination (c-index)*: OS c-index 0.76 (95% CI 0.65, 0.80); PFS: c-index 0.74 (95% CI 0.70, 0.82) *Model calibration*: NR *Other model performance measures*: NR
Rios-Velazquez 2014 *External validation (VUMC cohort)*		*Model discrimination (c-index)*: OS: c-index 0.72 (95% CI 0.64, 0.78); PFS: c-index 0.64 (95% CI 0.59, 0.72) *Model calibration*: NR *Other model performance measures*: NR
Rietbergen 2013, the Netherlands	**Rietbergen 2013 model** HPV status: positive vs negative	*Model discrimination (c-index)*: OS: c-index 0.68 (95% CI 0.65-0.71) *Model calibration*: NR

Development	Comorbidity: ACE 0 vs ACE 1-2 vs ACE 3 (HPV-); ACE 0-1 vs ACE 2-3 (HPV+) Nodal stage (in HPV-): N0-N2a vs N2b-N3 Tumour stage (in HPV+ with ACE 0): T1-2 vs T3-4 *Flowchart presented on how to assign low, intermediate and high risk.*	*Other model performance measures*: NR
Rietbergen 2015 *External validation*		*Model discrimination (c-index)*: 5 yr OS: c-index 0.69 (95% CI 0.63, 0.75); 5 yr PFS: c-index 0.66 (95% CI 0.59-0.74) *Model calibration*: NR *Other model performance measures*: NR
Deschuymer 2018, Belgium *External validation*		*Model discrimination (c-index)*: OS: c-index 0.58 (0.53, 0.64); OS: c-index HPV+ group only: 0.62 (0.52, 0.73); LRC, DMC: NR *Model calibration*: NR *Other model performance measures*: NR
Huang 2015, Canada *Development*	**Huang 2015 RPA model** Recursive partitioning analysis model RPA-I: T1-3N0-2b RPA-II: T1-3N2c RPA-III: T4 or N3	*Model discrimination (c-index)*: NR *Model calibration*: NR *Other model performance measures*: Based on an overall score derived from the hazard consistency score, the hazard discrimination score, explained variance score, hazard difference score and sample size balance score, the RPA risk prediction model performed best, followed by the AHR model and then the AJCC/UICC 7th edition. *(NB Further model developed based on RPA stage, age and smoking developed in Huang 2015 but not validated)*
Keane 2016, US *External validation*		*Model discrimination (c-index)*: OS c-index 0.60 (95% CI 0.59, 0.61); HNC-specific mortality: c-index 0.62 (95% CI 0.61, 0.63); AJCC staging system: OS: c-index 0.54 (95% CI 0.53, 0.55); HNC-specific mortality: c-index 0.55 (95% CI 0.54, 0.57) *Model calibration*: NR *Other model performance measures*: NR
Huang 2015, Canada *Development*	**Huang 2015 AHR model** Adjusted hazard ratio derived model AHR-I: T1N0-N2b or T2N0-N2a AHR-II: T1N2c, T2N2b-N2c, or T3N0-N2b AHR-III: T1-2N3, T3N2c, or T4N0-N2a AHR-IVA: T3N3 or T4N2b-N3 AHR-IVB: M1	*Model discrimination (c-index)*: NR *Model calibration*: NR *Other model performance measures*: Based on an overall score derived from the hazard consistency score, the hazard discrimination score, explained variance score, hazard difference score and sample size balance score, the RPA risk prediction model performed best, followed by the AHR model and then the AJCC/UICC 7th edition. *(NB Further model based on RPA stage, age and smoking developed in Huang 2015 but not validated)*
O'Sullivan 2016, Canada *External validation (in both training cohort and validation cohort)*		*Model discrimination (c-index)*: NR *Model calibration*: NR *Other model performance measures*: Based on an overall score derived from the hazard consistency score, the hazard discrimination score, explained variance score, hazard difference score and sample size balance score, the AHR risk classification performed best, followed by the RPA model and then the AJCC/UICC 7th edition.

O'Sullivan 2016, Canada *Development*	**O'Sullivan 2016 AHR model** Adjusted hazard ratio derived model AHR-New I: T1–T2N0–N2b AHR-New II: T1–T2N2c or T3N0–N2c AHR-New III: T4 or N3 **O'Sullivan 2016 RPA model** RPA stage I: T1–T3N0–N2b RPA stage II: T1–T3N2c RPA stage III: T4 or N3	*Model discrimination (c-index)*: NR *Model calibration*: NR *Other model performance measures*: Based on an overall score derived from the hazard consistency score, the hazard discrimination score, explained variance score, hazard difference score and sample size balance score, the AHR (NEW) risk classification performed best (and as well as the AHR Original), followed by the RPA model and then the AJCC/UICC 7th edition.
O'Sullivan 2016, Canada *External validation*		*Model discrimination (c-index)*: NR *Model calibration*: NR *Other model performance measures*: Based on an overall score derived from the hazard consistency score, the hazard discrimination score, explained variance score, hazard difference score and sample size balance score, the AHR (NEW) risk classification performed best, followed by AHR Original, the RPA model and then the AJCC/UICC 7th edition.
Alabi 2022 *Development*	**ProgTOOL** Age Sex: 0=male, 1=female Ethnicity: 0=white, 1=black, 2=other Marital status: 1=married, 0=single Tumour grade: I-IV HPV status: 0=negative, 1=positive Site: 1=base of tongue, 2=oropharynx, 3=tonsils 4=n=vallecular T-stage (AJCC 7th ed): 1-4 N=stage (AJCC 7th ed): 1-3 M-stage (AJCC 7th ed): 0-1 Surgery: 0=no, 1=yes Surgery +radiotherapy: 0=no, 1=yes Chemotherapy: 0=no, 1=yes Disease free survival (months): 0-500 http://oncotelligence.com/	*Model discrimination (c-index)*: NR *Model calibration*: NR *Other model performance measures:* PPV: 0.97, NPV: 0.76, sensitivity: 0.89, specificity: 0.92, F1 score: 0.93, Accuracy: 89.9%, balanced accuracy: 86.3%, weighted accuracy: 92.5%, Matthews' correlation coefficient: 0.77, weighted AUC: 0.929
Alabi 2023 *External validation*		*Model discrimination (c-index)*: NR *Model calibration*: Slope not fitted so difficult to interpret. Accuracy of prediction appears to vary depending on lower/higher predicted probabilities. *Other model performance measures:* PPV: 0.93, NPV: 0.89, sensitivity: 0.76, specificity: 0.97, F1 score: 0.84, Accuracy: 90.2%, balanced accuracy: 86.5%, Matthews' correlation coefficient: 0.78, weighted AUC: 0.94, Brier score 0.06, Net Benefit value of model approximately 0.7 at 10% - 50% probability threshold.

Risk of bias summary

Figure: PROBAST summary Chart shows percentage of study cohorts meeting/not meeting criteria: Y=yes; PY=probably yes; NI=no or insufficient information; PN=probably no; N=no. AS=all study cohorts; MD=model development cohorts; EV=external validation cohorts. Numbers of cohorts contributing to the different criteria varies (e.g. as not all evaluations reporting both OS and PFS; the criterion *'participants with missing data handled appropriately'* is only applicable where there was missing data). Every evaluation counted for the analysis domain; some cohorts were used for evaluating more than one model. The criterion *'all enrolled participants included in analysis'* was answered with 'no' if participants were excluded on the basis of missing variable data.